PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/82, 9/02, 5/10, A01H 5/00

A2

(11) International Publication Number: WO 99/55887

(43) International Publication Date: 4 November 1999 (04.11.99)

(21) International Application Number: PCT/US99/08384

(22) International Filing Date: 16 April 1999 (16.04.99)

(30) Priority Data:
60/083,042 24 April 1998 (24.04.98) US

(71) Applicant *(for all designated States except US)*: E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US). KINNEY, Anthony, J. [GB/US]; 609 Lore Avenue, Wilmington, DE 19809 (US). PEARLSTEIN, Richard, W. [US/US]; 12 Canoe Court, Newark, DE 19702 (US). WILLIAMS, Mark, E. [US/US]; 37 Country Hills Drive, Newark, DE 19711 (US).

(74) Agent: MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AE, AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

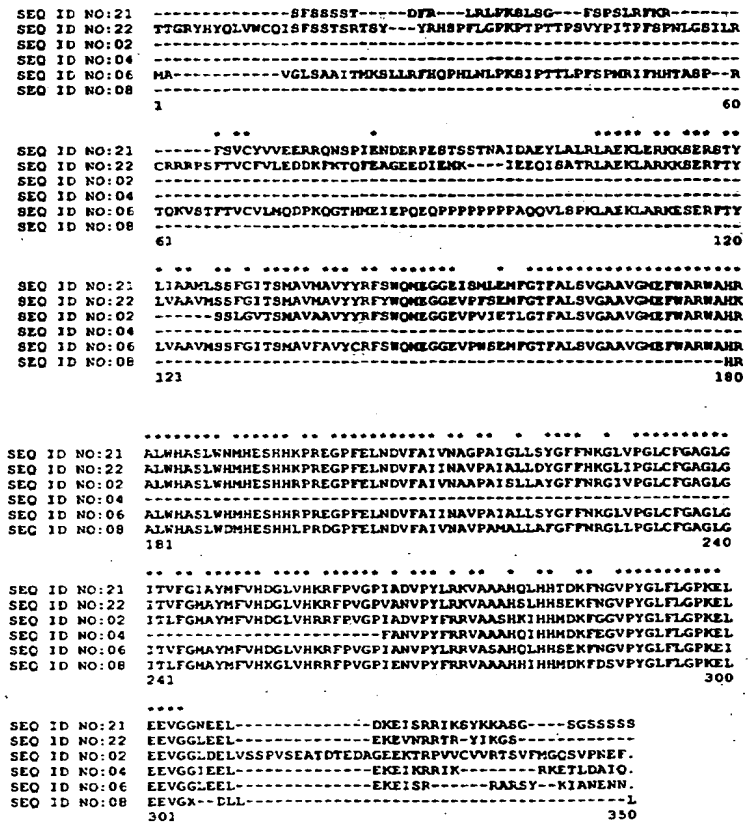
Published

Without international search report and to be republished upon receipt of that report.

(54) Title: CAROTENOID BIOSYNTHESIS ENZYMES

(57) Abstract

This invention relates to an isolated nucleic acid fragment encoding a carotenoid biosynthetic enzyme. The invention also relates to the construction of a chimeric gene encoding all or a portion of the carotenoid biosynthetic enzyme, is sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the carotenoid biosynthetic enzyme in a transformed host cell.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

TITLE

CAROTENOID BIOSYNTHESIS ENZYMES

This application claims the benefit of U.S. Provisional Application No. 60/083,042, filed April 24, 1998.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding enzymes of the carotenoid biosynthesis pathway in plants and seeds.

BACKGROUND OF THE INVENTION

Plant carotenoids are orange and red lipid-soluble pigments found embedded in the membranes of chloroplasts and chromoplasts. In leaves and immature fruits the color is masked by chlorophyll but in later stages of development these pigments contribute to the bright color of flowers and fruits. Carotenoids protect against photoxidation processes and harvest light for photosynthesis. The carotenoid biosynthesis pathway leads to the production of abscisic acid with intermediaries useful in the agricultural and food industries as well as products thought to be involved in cancer prevention (Bartley, G. E. and Scolnik, P. A. (1995) *Plant Cell* 7: 1027-1038).

One of the intermediaries in the carotenoid biosynthesis pathway, lycopene, may have one of two different fates: through the action of lycopene epsilon cyclase it may become alpha carotene, or it may be transformed into beta carotene by lycopene cyclase. Beta-carotene dehydroxylase converts beta-carotene into zeaxanthin. Zeaxanthin epoxidase transforms zeaxanthin into violxanthin and eventually absisic acid.

Zeaxanthin is the bright orange product highly prized as a pigmenting agent for animal feed which makes the meat fat, skin, and egg yolks a dark yellow (Scott, M. L. et al. (1968) *Poultry Sci.* 47:863-872). Gram per gram, zeaxanthin is one of the best pigmenting compounds because it is highly absorbable. Yellow corn, which produces one of the best ratios of lutein to zeaxanthin contains in average 20 to 25 mg of xanthophyll per kg while marigold petals yield 6,000 to 10,000 mg/kg.

Enzymes from the carotenoid pathway have previously been isolated from a variety of bacteria, fungi and higher plants. There is a great variety in the functions and properties of the enzymes emanating from different sources and the amino acid similarities between bacterial and plant enzymes is very low. Therefore, knowing the amino acid sequence of the enzyme from a bacteria or a plant will not necessarily make it easy to screen for the enzyme in another source.

SUMMARY OF THE INVENTION

The instant invention relates to isolated nucleic acid fragments encoding carotenoid biosynthetic enzymes. Specifically, this invention concerns an isolated nucleic acid fragment encoding a beta carotene hydroxylase, a lycopene cyclase or a lycopene epsilon cyclase. In addition, this invention relates to a nucleic acid fragment that is complementary

to the nucleic acid fragment encoding beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase.

An additional embodiment of the instant invention pertains to a polypeptide encoding all or a substantial portion of a carotenoid biosynthetic enzyme selected from the group consisting of beta carotene hydroxylase, lycopene cyclase and lycopene epsilon cyclase.

In another embodiment, the instant invention relates to a chimeric gene encoding a beta carotene hydroxylase, a lycopene cyclase or a lycopene epsilon cyclase, or to a chimeric gene that comprises a nucleic acid fragment that is complementary to a nucleic acid fragment encoding a beta carotene hydroxylase, a lycopene cyclase or a lycopene epsilon cyclase, operably linked to suitable regulatory sequences, wherein expression of the chimeric gene results in production of levels of the encoded protein in a transformed host cell that is altered (i.e., increased or decreased) from the level produced in an untransformed host cell.

In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding a beta carotene hydroxylase, a lycopene cyclase or a lycopene epsilon cyclase, operably linked to suitable regulatory sequences. Expression of the chimeric gene results in production of altered levels of the encoded protein in the transformed host cell. The transformed host cell can be of eukaryotic or prokaryotic origin, and include cells derived from higher plants and microorganisms. The invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

An additional embodiment of the instant invention concerns a method of altering the level of expression of a beta carotene hydroxylase, a lycopene cyclase or a lycopene epsilon cyclase in a transformed host cell comprising: a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a beta carotene hydroxylase, a lycopene cyclase or a lycopene epsilon cyclase; and b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase in the transformed host cell.

An addition embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or a substantial portion of an amino acid sequence encoding a beta carotene hydroxylase, a lycopene cyclase or a lycopene epsilon cyclase.

A further embodiment of the instant invention is a method for evaluating at least one compound for its ability to inhibit the activity of a beta carotene hydroxylase, a lycopene cyclase or a lycopene epsilon cyclase, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a beta carotene hydroxylase, a lycopene cyclase or a lycopene epsilon cyclase, operably linked to suitable regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results

in production of beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase in the transformed host cell: (c) optionally purifying the beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase expressed by the transformed host cell; (d) treating the beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase with a compound to be tested; and (e) comparing the activity of the beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase that has been treated with a test compound to the activity of an untreated beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase, thereby selecting compounds with potential for inhibitory activity.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

Figure 1 depicts the amino acid sequence alignment between the beta carotene hydroxylase from a corn contig assembled from clones csi1.pk0019.g1 and csi1n.pk0035.f7 (SEQ ID NO:2), a rice contig assembled from clones rl0n.pk102.k19 and rls72.pk0008.e3 (SEQ ID NO:4), soybean clone sfl1.pk0049.g8 (SEQ ID NO:6), wheat clone wkm2n.pk008.a5 (SEQ ID NO:8), and *Arabidopsis thaliana* (NCBI gi Accession No. 1575296, SEQ ID NO:21) and *Capsicum annuum* (NCBI gi Accession No. 2956717, SEQ ID NO:22) beta carotene hydroxylases. Amino acids which are conserved among all sequences are indicated with an asterisk (*). Dashes are used by the program to maximize alignment of the sequences.

Figure 2 depicts the amino acid sequence alignment between the lycopene cylcase from a corn contig assembled from clones p0110.cgsmy23r, p0110.cgsmj48r and cr1n.pk0051.h6 (SEQ ID NO:10), soybean clone sfl1.pk0034.c1 (SEQ ID NO:12), wheat clone wle1n.pk0059.f5 (SEQ ID NO:14) and *Arabidopsis thaliana* (NCBI gi Accession No. 735882, SEQ ID NO:23). Amino acids which are conserved among all sequences are indicated with an asterisk (*). Dashes are used by the program to maximize alignment of the sequences.

Figure 3 depicts the amino acid sequence alignment between the lycopene epsilon cyclase from corn clone cen3n.pk0135.f10 (SEQ ID NO:16), a corn contig assembled from clones p0126.cnldl94r and p0126.cnldj12r (SEQ ID NO:18), soybean clone sr1.pk0068.b1 (SEQ ID NO:20) and *Lycopersicon esculentum* (NCBI gi Accession No. 3005983, SEQ ID NO:24). Dashes are used by the program to maximize alignment of the sequences.

The following sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

SEQ ID NO:1 is the nucleotide sequence comprising the contig assembled from the entire cDNA insert in clone csi1n.pk0035.f7 and a portion of the cDNA insert in clone csi1.pk0019.g1 encoding the C-terminal three quarters of a corn beta carotene hydroxylase.

SEQ ID NO:2 is the deduced amino acid sequence of the C-terminal three quarters of a corn beta carotene hydroxylas derived from the nucleotide sequence of SEQ ID NO:1.

SEQ ID NO:3 is the nucleotide sequence comprising the contig assembled from the entire cDNA insert in clone rls72.pk0008.e3 and a portion of the cDNA insert in clone rl0n.pk102.k19 encoding a portion of a rice beta carotene hydroxylase.

SEQ ID NO:4 is the deduced amino acid sequence of a portion of a rice beta carotene hydroxylase derived from the nucleotide sequence of SEQ ID NO:3.

SEQ ID NO:5 is the nucleotide sequence comprising the entire cDNA insert in clone sfl1.pk0049.g8 encoding an entire soybean beta carotene hydroxylase.

SEQ ID NO:6 is the deduced amino acid sequence of an entire soybean beta carotene hydroxylase derived from the nucleotide sequence of SEQ ID NO:5.

SEQ ID NO:7 is the nucleotide sequence comprising a portion of the cDNA insert in clone wkm2n.pk008.a5 encoding the C-terminal half of a wheat beta carotene hydroxylase.

SEQ ID NO:8 is the deduced amino acid sequence of the C-terminal half of a wheat beta carotene hydroxylase derived from the nucleotide sequence of SEQ ID NO:9.

SEQ ID NO:9 is the nucleotide sequence comprising the contig assmebled from the entire cDNA insert in clone cr1n.pk0051.h6 and a portion of the cDNA insert in clones p0110.cgsmy23r and p0110.cgsmj48r encoding a substantial portion of a corn lycopene cyclase.

SEQ ID NO:10 is the deduced amino acid sequence of a substantial portion of a corn lycopene cyclase derived from the nucleotide sequence of SEQ ID NO:9.

SEQ ID NO:11 is the nucleotide sequence comprising the entire cDNA insert in clone sfl1.pk0034.c1 encoding an entire soybean lycopene cyclase.

SEQ ID NO:12 is the deduced amino acid sequence of an entire soybean lycopene cyclase derived from the nucleotide sequence of SEQ ID NO:11.

SEQ ID NO:13 is the nucleotide sequence comprising the entire cDNA insert in clone wle1n.pk0059.f5 encoding a portion of a wheat lycopene cyclase.

SEQ ID NO:14 is the deduced amino acid sequence of a portion of a wheat lycopene cyclase derived from the nucleotide sequence of SEQ ID NO:13.

SEQ ID NO:15 is the nucleotide sequence comprising the entire cDNA insert in clone cen3n.pk0135.f10 encoding a portion of a corn lycopene epsilon cyclase.

SEQ ID NO:16 is the deduced amino acid sequence of a portion of a corn lycopene epsilon cyclase derived from the nucleotide sequence of SEQ ID NO:15.

SEQ ID NO:17 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA insert in clones p0126.cnldl94r and p0126.cnldj12r encoding a portion of a corn lycopene epsilon cyclase.

SEQ ID NO:18 is the deduced amino acid sequence of a portion of a corn lycopene epsilon cyclase derived from the nucleotide sequence of SEQ ID NO:17.

SEQ ID NO:19 is the nucleotide sequence comprising the entire cDNA insert in clone sr1.pk0068.b1 encoding the C-terminal two thirds of a soybean lycopene epsilon cyclase.

SEQ ID NO:20 is the deduced amino acid sequence of the C-terminal two thirds of a soybean lycopene epsilon cyclase derived from the nucleotide sequence of SEQ ID NO:19.

SEQ ID NO:21 is the amino acid sequence of an *Arabidopsis thaliana* beta carotene hydroxylase, NCBI gi Accession No. 1575296.

SEQ ID NO:22 is the amino acid sequence of a *Capsicum annuum* beta carotene hydroxylase NCBI gi Accession No. 2956717.

SEQ ID NO:23 is the amino acid sequence of an *Arabidopsis thaliana* lycopene cyclase NCBI gi Accession No. 735882.

SEQ ID NO:24 is the amino acid sequence of a *Lycopersicon esculentum* lycopene epsilon cyclase, NCBI gi Accession No. 3005983.

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research 13*:3021-3030 (1985) and in the *Biochemical Journal 219 (No. 2)*:345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA. As used herein, "contig" refers to an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide sequence. For example, several DNA sequences can be compared and aligned to identify common or overlapping regions. The individual sequences can then be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional

properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded protein, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, the skilled artisan recognizes that substantially similar nucleic acid sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the sequences exemplified herein. Preferred substantially similar nucleic acid fragments of the instant invention are those nucleic acid fragments whose DNA sequences are 80% identical to the coding sequence of the nucleic acid fragments reported herein. More preferred nucleic acid fragments are 90% identical to the coding sequence of the nucleic acid fragments reported herein. Most preferred are nucleic acid fragments that are 95% identical to the coding sequence of the nucleic acid fragments reported herein. Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino

acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment that encodes all or a substantial portion of the amino acid sequence encoding the beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase proteins as set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18 and 20. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) *Biochemistry of Plants 15*:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The

translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G. D. (1995) *Molecular Biotechnology 3*:225).

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., (1989) *Plant Cell 1*:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, J. J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys. 100*:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include Agrobacterium-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol. 143*:277) and particle-accelerated or "gene gun" transformation technology (Klein T. M. et al. (1987) *Nature (London) 327*:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several carotenoid biosynthetic enzymes have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. Table 1 lists the proteins that are described herein, and the designation of the cDNA clones that comprise the nucleic acid fragments encoding these proteins.

TABLE 1

Carotenoid Biosynthetic Enzymes

Enzyme	Clone	Plant
Beta Carotene Hydroxylase	Contig of: csi1.pk0019.g1 csi1n.pk0035.f7	Corn
	Contig of: rl0n.pk102.k19 rls72.pk0008.e3	Rice
	sfl1.pk0049.g8	Soybean
	wkm2n.pk008.a5	Wheat
Lycopene Cyclase	Contig of: p0110.cgsmy23r p0110.cgsmj48r cr1n.pk0051.h6	corn
	sfl1.pk0034.c1	Soybean
	wle1n.pk0059.f5	Wheat
Lycopene Epsilon Cyclase	cen3n.pk0135.f10	Corn
	Contig of: p0126.cnldl94r p0126.cnldj12r	Corn
	sr1.pk0068.b1	Soybean

The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after

amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., (1988) *Proc. Natl. Acad. Sci. USA 85*:8998) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., (1989) *Proc. Natl. Acad. Sci. USA 86*:5673; Loh et al., (1989) *Science 243*:217). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman, M. A. and Martin, G. R., (1989) *Techniques 1*:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner, R. A. (1984) *Adv. Immunol. 36*:1: Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of alpha carotene or beta carotene in those cells.

Co-suppression of the lycopene epsilon cyclase in corn endosperm should divert more of the xanthophill biosynthesis towards zeaxanthin. Zeaxanthin gives a consumer-preferred darker color to yolks and poultry products. The xanthophyll present in maize is well utilized by poultry but is present in very small amounts (25 to 50 mg/Kg compared to 6,000 to 10,000 in marigolds). Thus, increasing the amount of zeaxanthin in corn would produce a more desirable product for poultry feed. Hydroxylases catalyze the last step in xanthophyll biosynthesis. Increasing hydroxylase activity in corn endosperm may increase xanthophyll content. Blocking hydroxylase activity may create a high bata-carotene corn which may be valuable for human consumption as well as animal feed.

Overexpression of the beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) *EMBO J. 4*:2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics 218*:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant carotenoid biosynthetic enzyme to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by altering the coding sequence to encode beta carotene hydroxylase with appropriate intracellular targeting sequences such as transit sequences (Keegstra, K. (1989) *Cell 56*:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels, J. J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53), or nuclear localization signals (Raikhel, N. (1992) *Plant Phys. 100*:1627-1632) added and/or with targeting sequences that are already present removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant carotenoid biosynthetic enzyme can be constructed by linking a gene or gene fragment encoding a beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via

transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

The instant beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded carotenoid biosynthetic enzyme. An example of a vector for high level expression of the instant beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase in a bacterial host is provided (Example 8).

Additionally, the instant beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase can be used as targets to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because the beta carotene hydroxylase, lycopene cyclase and lycopene epsilon cyclase described herein catalyze various steps in carotenoid biosynthesis. Accordingly, inhibition of the activity of one or more of the enzymes described herein could lead to inhibition plant growth. Thus, the instant beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase could be appropriate for new herbicide discovery and design.

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et at., (1987) *Genomics 1*:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map

previously obtained using this population (Botstein, D. et al., (1980) *Am. J. Hum. Genet. 32*:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in R. Bernatzky, R. and Tanksley, S. D. (1986) *Plant Mol. Biol. Reporter 4(1)*:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; *see* Hoheisel, J. D., et al., In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask, B. J. (1991) *Trends Genet. 7*:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan, M. et al. (1995) *Genome Research 5*:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian, H. H. (1989) *J. Lab. Clin. Med. 114(2)*:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield, V. C. et al. (1993) *Genomics 16*:325-332), allele-specific ligation (Landegren, U. et al. (1988) *Science 241*:1077-1080), nucleotide extension reactions (Sokolov, B. P. (1990) *Nucleic Acid Res. 18*:3671), Radiation Hybrid Mapping (Walter, M. A. et al. (1997) *Nature Genetics 7*:22-28) and Happy Mapping (Dear, P. H. and Cook, P. R. (1989) *Nucleic Acid Res. 17*:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer, (1989) *Proc. Natl. Acad. Sci USA 86*:9402; Koes et al., (1995) *Proc. Natl. Acad. Sci USA 92*:8149; Bensen et al., (1995) *Plant Cell 7*:75). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may

be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding a beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the beta carotene hydroxylase, lycopene cyclase or lycopene epsilon cyclase gene product.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries: Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean and Wheat

Library	Tissue	Clone
cen3n	Corn Endosperm 20 Days After Pollination*	cen3n.pk0135.f10
cr1n	Corn Root From 7 Day Old Seedlings*	cr1n.pk0051.h6
csi1	Corn Silk	csi1.pk0019.g1
csi1n	Corn Silk*	csi1n.pk0035.f7
p0110	Corn Leaf Tissue (minus midrib), Stages V3/V4 Infiltrated With Salicylic Acid for 4 Hours, 24 Hours and 7 Days, Pooled*	p0110.cgsmj48r p0110.cgsmy23r
p0126	Corn Leaf Tissue From V8-V10 Stages, Pooled, Night-Harvested	p0126.cnldj12r p0126.cnldl94r

Library	Tissue	Clone
rl0n	Rice 15 Day Old Leaf*	rl0n.pk102.k19
rls72	Rice Leaf 15 Days After Germination, 72 Hours After Infection of Strain *Magaporthe grisea* 4360-R-67 (AVR2-YAMO); Susceptible	rls72.pk0008.e3
sfl1	Soybean Immature Flower	sfl1.pk0034.c1
		sfl1.pk0049.g8
sr1	Soybean Root	sr1.pk0068.b1
wkm2n	Wheat Kernel Malted 175 Hours at 4 Degrees Celsius*	wkm2n.pk008.a5
wle1n	Wheat Leaf From 7 Day Old Etiolated Seedling*	wle1n.pk0059.f5

*These libraries were normalized essentially as described in U.S. Patent No. 5,482,845

cDNA libraries were prepared in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). Conversion of the Uni-ZAP™ XR libraries into plasmid libraries was accomplished according to the protocol provided by Stratagene. Upon conversion, cDNA inserts were contained in the plasmid vector pBluescript. cDNA inserts from randomly picked bacterial colonies containing recombinant pBluescript plasmids were amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences or plasmid DNA was prepared from cultured bacterial cells. Amplified insert DNAs or plasmid DNAs were sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams, M. D. et al., (1991) *Science 252*:1651). The resulting ESTs were analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

ESTs encoding carotenoid biosynthetic enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. (1993) *Nature Genetics 3*:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA

sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding Beta Carotene Hydroxylase

The BLASTX search using the EST sequences from clones csi1.pk0019.g1, csi1n.pk0035.f7, rls72.pk0008.e3, and a contig formed by the sequences from clones ssm.pk0030.d2, sfl1.pk0049.g8, se2.12e09, ssm.pk0023.b3, and ssm.pk0030.d2 revealed similarity of the proteins encoded by the cDNAs to Beta Carotene Hydroxylase from *Arabidopsis thaliana* (GenBank Accession No. U58919). The BLAST results for each of these ESTs and the contig are shown in Table 3:

TABLE 3

BLAST Results for Clones Encoding Polypeptides Homologous to Beta Carotene Hydroxylase

Clone	BLAST pLog Score U58919
csi1.pk0019.g1	64.26
csi1n.pk0035.f7	27.10
rls72.pk0008.e3	7.53
Contig formed of ssm.pk0030.d2 sfl1.pk0049.g8 se2.12e09 ssm.pk0023.b3 ssm.pk0030.d2	27.39

TBLASTN analysis of the proprietary plant EST database indicated that a wheat clone also encoded beta carotene hydroxylase. The BLASTX search using the EST sequences from clone wkm2n.pk008.a5 revealed similarity of the proteins encoded by the cDNAs to beta carotene hydroxylase 2 from *Capsicum annuum* (NCBI General Identifier No. 2956717), with a pLog value of 72.22.

The sequence of the entire cDNA insert in clone csi1n.pk0035.f7 was determined and a contig assembled with this sequence and the EST sequence from clone csi1.pk0019.g1. The sequence from this contig is shown in SEQ ID NO:1; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:2. The sequence of the entire cDNA insert in clone rls72.pk0008.e3 was determined and a contig assembled with this sequence and the EST sequence from clone rl0n.pk102.k19. The sequence from this contig is shown in SEQ ID NO:3; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:4. The sequence of the entire cDNA insert in clone sfl1.pk0049.g8 was determined and is shown in

SEQ ID NO:5; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:6. The amino acid sequence set forth in SEQ ID NO:6 was evaluated by BLASTP, yielding a pLog value of 122 versus the *Capsicum annuum* sequence (NCBI General Identifier No. 2956717; SEQ ID NO:24). The sequence of a portion of the cDNA insert from clone wkm2n.pk008.a5 is shown in SEQ ID NO:7; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:8. Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6 and 8 with the *Arabidopsis thaliana* and the *Capsicum annuum* sequences (SEQ ID NO:21 and SEQ ID NO:22). The data in Table 4 presents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6 and 8 and the *Capsicum annuum* sequence.

TABLE 4

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Beta Carotene Hydroxylase

Clone	SEQ ID NO.	Percent Identity to 2956717
Contig of: csi1.pk0019.g1 csi1n.pk0035.f7	2	67.9
Contig of: rl0n.pk102.k19 rls72.pk0008.e3	4	64.1
sfl1.pk0049.g8	6	66.3
wkm2n.pk008.a5	8	78.6

Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the amino acid sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS.* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode entire soybean beta carotene hydroxylase and portions of corn, rice and wheat beta carotene hydroxylase. These sequences represent the first monocot and the first soybean sequences encoding beta carotene hydroxylase.

EXAMPLE 4

Characterization of cDNA Clones Encoding Lycopene Cyclase

The BLASTX search using the EST sequences from clones cr1n.pk0051.h6 and wre1n.pk0146.g4 revealed similarity of the proteins encoded by the cDNAs to Lycopene Cyclase from *Arabidopsis thaliana* (GenBank Accession No. L40176). The BLAST results for each of these ESTs are shown in Table 5:

TABLE 5

BLAST Results for Clones Encoding Polypeptides Homologous to Lycopene Cyclase

Clone	BLAST pLog Score GenBank Accession No. L40176
cr1n.pk0051.h6	19.52
wre1n.pk0146.g4	28.00

TBLASTN analysis of the proprietary plant EST database indicated that a soybean clone also encoded lycopene cyclase. The BLASTX search using the EST sequence from clone sfl1.pk0034.c1 revealed similarity of the proteins encoded by the cDNAs to capsanthin-capsorubin synthase from *Capsicum annuum* (NCBI General Identifier No. 626009), with a pLog value of >254. The *Capsicum annuum* capsanthin-capsorubin synthase is a chromoplast lycopene cyclase (Hugueney, P. et al (1995) *Plant J. 8*:417-424)..

The sequence of the entire cDNA insert in clone cr1n.pk0051.h6 was determined and a contig assembled with this sequence and the EST sequences from clones p0110.cgsmy23r and p0110.cgsmj48r. This contig sequence is shown in SEQ ID NO:9; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:10. The amino acid sequence set forth in SEQ ID NO:10 was evaluated by BLASTP, yielding a pLog value of 153 versus the *Arabidopsis thaliana* sequence (NCBI General Identifier No. 735882; SEQ ID NO:23). The sequence of the entire cDNA insert in clone sfl1.pk0034.c1 was determined and is shown in SEQ ID NO:11; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:12. The amino acid sequence set forth in SEQ ID NO:12 was evaluated by BLASTP, yielding a pLog value of >254 versus the *Capsicum annuum* sequence (NCBI General Identifier No. 626009). The sequence of the entire cDNA insert in clone wle1n.pk0059.f5 was determined and is shown in SEQ ID NO:13; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:14. The amino acid sequence set forth in SEQ ID NO:14 was evaluated by BLASTP, yielding a pLog value of 82.52 versus the *Arabidopsis thaliana* sequence (NCBI General Identifier No. 735882). Figure 2 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:10, 12 and 14 and the *Arabidopsis thaliana* sequence (SEQ ID NO:23). The data in Table 6 presents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:10, 12 and 14 and the *Arabidopsis thaliana* sequence.

TABLE 6

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Lycopene Cyclase

Clone	SEQ ID NO.	Percent Identity to 735882
Contig of: p0110.cgsmy23r p0110.cgsmj48r cr1n.pk0051.h6	10	72.3
sfl1.pk0034.c1	12	50.4
wle1n.pk0059.f5	14	67.6

Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the amino acid sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS.* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode entire soybean lycopene cylcase and portions of corn and wheat lycopene cyclase. These sequences represent the first monocot and the first soybean sequences encoding lycopene cyclase.

EXAMPLE 5

Characterization of cDNA Clones Encoding Lycopene Epsilon Cyclase

The BLASTX search using the nucleotide sequence from clone cen3n.pk0135.f10, and the EST sequence from clone sr1.pk0068.b1 revealed similarity of the proteins encoded by the cDNAs to Lycopene Epsilon Cyclase from *Arabidopsis thaliana* (GenBank Accession No. U50738). The BLAST results for each of these ESTs are shown in Table 7:

TABLE 7

BLAST Results for Clones Encoding Polypeptides Homologous to Lycopene Epsilon Cyclase

Clone	BLAST pLog Score GenBank Accession No. U50738
cen3n.pk0135.f10	37.70
sr1.pk0068.b1	31.40

TBLASTN analysis of the proprietary plant EST database indicated that other corn clones besides cen3n.pk0135.f10 also encoded lycopene epsilon cyclase. The BLASTX search using the contig assembled of the EST sequences from clones p0126.cnldl94r and p0126.cnldj12rrevealed similarity of the proteins encoded by the cDNAs to lycopene epsilon

cyclase from *Lycopersicon esculentum* (NCBI General Identifier No. 3005983), with a pLog value of 75.00.

The nucleotide sequence of the entire cDNA insert in clone cen3n.pk0135.f10 was determined and is shown in SEQ ID NO:15; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:16. The amino acid sequence set forth in SEQ ID NO:16 was evaluated by BLASTP, yielding a pLog value of 40.70 versus the *Lycopersicon esculentum* sequence. The nucleotide sequence of the contig assembled from a portion of the cDNA insert in clones p0126.cnldl94r and p0126.cnldj12r is shown in SEQ ID NO:17; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:17. The sequence of the entire cDNA insert in clone sr1.pk0068.b1 was determined and is shown in SEQ ID NO:19; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:20. The amino acid sequence set forth in SEQ ID NO:20 was evaluated by BLASTP, yielding a pLog value of 127.0 vresus the *Lycopersicon esculentum* sequence.

Figure 3 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:16, 18 and 20 and the *Lycopersicon esculentum* sequence (SEQ ID NO:24). The data in Table 8 presents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:16, 18 and 20 and the *Lycopersicon esculentum* sequence.

TABLE 8

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Lycopene Epsilon Cyclase

Clone	SEQ ID NO.	Percent Identity to 3005983
cen3n.pk0135.f10	16	71.1
Contig of: p0126.cnldl94r p0126.cnldj12r	18	63.2
sr1.pk0068.b1	20	68.6

Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the amino acid sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode portions of corn lycopene epsilon cyclase and a nearly entire soybean lycopene epsilon cyclase. These sequences represent the first corn and soybean sequences encoding lycopene epsilon cyclase.

EXAMPLE 6

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding a carotenoid biosynthetic enzyme in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (Nco I or Sma I) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes Nco I and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb Nco I-Sma I fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb Sal I-Nco I promoter fragment of the maize 27 kD zein gene and a 0.96 kb Sma I-Sal I fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding a carotenoid biosynthetic enzyme, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al., (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers

resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein T. M. et al., (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 μm in diameter) are coated with DNA using the following technique. Ten μg of plasmid DNAs are added to 50 μL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 μL of a 2.5 M solution) and spermidine free base (20 μL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 μL of ethanol. An aliquot (5 μL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al., (1990) *Bio/Technology* *8*:833-839).

EXAMPLE 7

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem. 261*:9228-9238) can be used for expression of the instant carotenoid biosynthetic enzyme in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embroys may then be transformed with the expression vector comprising sequences encoding a carotenoid biosynthetic enzyme. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein T. M. et al. (1987) *Nature* (London) *327*:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature 313*:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene 25*:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the carotenoid

biosynthetic enzyme and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL $CaCl_2$ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 8

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant carotenoid biosynthetic enzymes can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene 56*:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 μg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 μL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 μg/mL ampicillin. Transformants containing the gene encoding the carotenoid biosynthetic enzyme are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol. 189*:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 μL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One μg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 9

Evaluating Compounds for Their Ability to Inhibit the Activity of Carotenoid Biosynthetic Enzymes

The carotenoid biosynthetic enzymes described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 8, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant carotenoid biosynthetic enzymes may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione

S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("$(His)_6$"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the enzyme.

Purification of the instant carotenoid biosynthetic enzymes, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the carotenoid biosynthetic enzymes are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, a carotenoid biosynthetic enzyme may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a $(His)_6$ peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include β–mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the carotenoid biosynthetic enzymes disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. For example, assays for beta carotene hydroxylase are presented by Sun Z. et al. (1996) *J. Biol. Chem. 271*:24349-24352. Assays for lycopene cyclase and lycopene epsilon cyclase are presented by Cunningham F. X. (1994) *Plant Cell 6*:1107-1121.

CLAIMS

What is claimed is:

1. An isolated nucleic acid fragment encoding all or a substantial portion of a lycopene epsilon cyclase comprising a member selected from the group consisting of:

(a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:16, SEQ ID NO:18 and SEQ ID NO:20;

(b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:16, SEQ ID NO:18 and SEQ ID NO:20; and

(c) an isolated nucleic acid fragment that is complementary to (a) or (b).

2. The isolated nucleic acid fragment of Claim 1 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:15, SEQ ID NO:17 and SEQ ID NO:19.

3. A chimeric gene comprising the nucleic acid fragment of Claim 1 operably linked to suitable regulatory sequences.

4. A transformed host cell comprising the chimeric gene of Claim 3.

5. A lycopene epsilon cyclase polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:16, SEQ ID NO:18 and SEQ ID NO:20.

6. An isolated nucleic acid fragment encoding all or a substantial portion of a beta carotene hydroxylase comprising a member selected from the group consisting of:

(a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8;

(b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8; and

(c) an isolated nucleic acid fragment that is complementary to (a) or (b).

7. The isolated nucleic acid fragment of Claim 6 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 and SEQ ID NO:7.

8. A chimeric gene comprising the nucleic acid fragment of Claim 6 operably linked to suitable regulatory sequences.

9. A transformed host cell comprising the chimeric gene of Claim 8.

10. A beta carotene hydroxylase polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8.

11. An isolated nucleic acid fragment encoding all or a substantial portion of a lycopene cyclase comprising a member selected from the group consisting of:

(a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14;

(b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14; and

(c) an isolated nucleic acid fragment that is complementary to (a) or (b).

12. The isolated nucleic acid fragment of Claim 11 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:9, SEQ ID NO:11 and SEQ ID NO:13.

13 A chimeric gene comprising the nucleic acid fragment of Claim 11 operably linked to suitable regulatory sequences.

14. A transformed host cell comprising the chimeric gene of Claim 13.

15. A lycopene cyclase polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:10, SEQ ID NO:12 and SEQ ID NO:14.

16. A method of altering the level of expression of a carotenoid biosynthetic enzyme in a host cell comprising:

(a) transforming a host cell with the chimeric gene of any of Claims 3, 8 and 13; and

(b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric gene

wherein expression of the chimeric gene results in production of altered levels of a carotenoid biosynthetic enzyme in the transformed host cell.

17. A method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a carotenoid biosynthetic enzyme comprising:

(a) probing a cDNA or genomic library with the nucleic acid fragment of any of Claims 1, 6 and 11;

(b) identifying a DNA clone that hybridizes with the nucleic acid fragment of any of Claims 1, 6 and 11;

(c) isolating the DNA clone identified in step (b); and

(d) sequencing the cDNA or genomic fragment that comprises the clone isolated in step (c)

wherein the sequenced nucleic acid fragment encodes all or a substantial portion of the amino acid sequence encoding a carotenoid biosynthetic enzyme.

18. A method of obtaining a nucleic acid fragment encoding a substantial portion of an amino acid sequence encoding a carotenoid biosynthetic enzyme comprising:

(a) synthesizing an oligonucleotide primer corresponding to a portion of the sequence set forth in any of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17 and 19 and

(b) amplifying a cDNA insert present in a cloning vector using the oligonucleotide primer of step (a) and a primer representing sequences of the cloning vector

wherein the amplified nucleic acid fragment encodes a substantial portion of an amino acid sequence encoding a carotenoid biosynthetic enzymes.

19. The product of the method of Claim 17.

20. The product of the method of Claim 18.

FIGURE 1

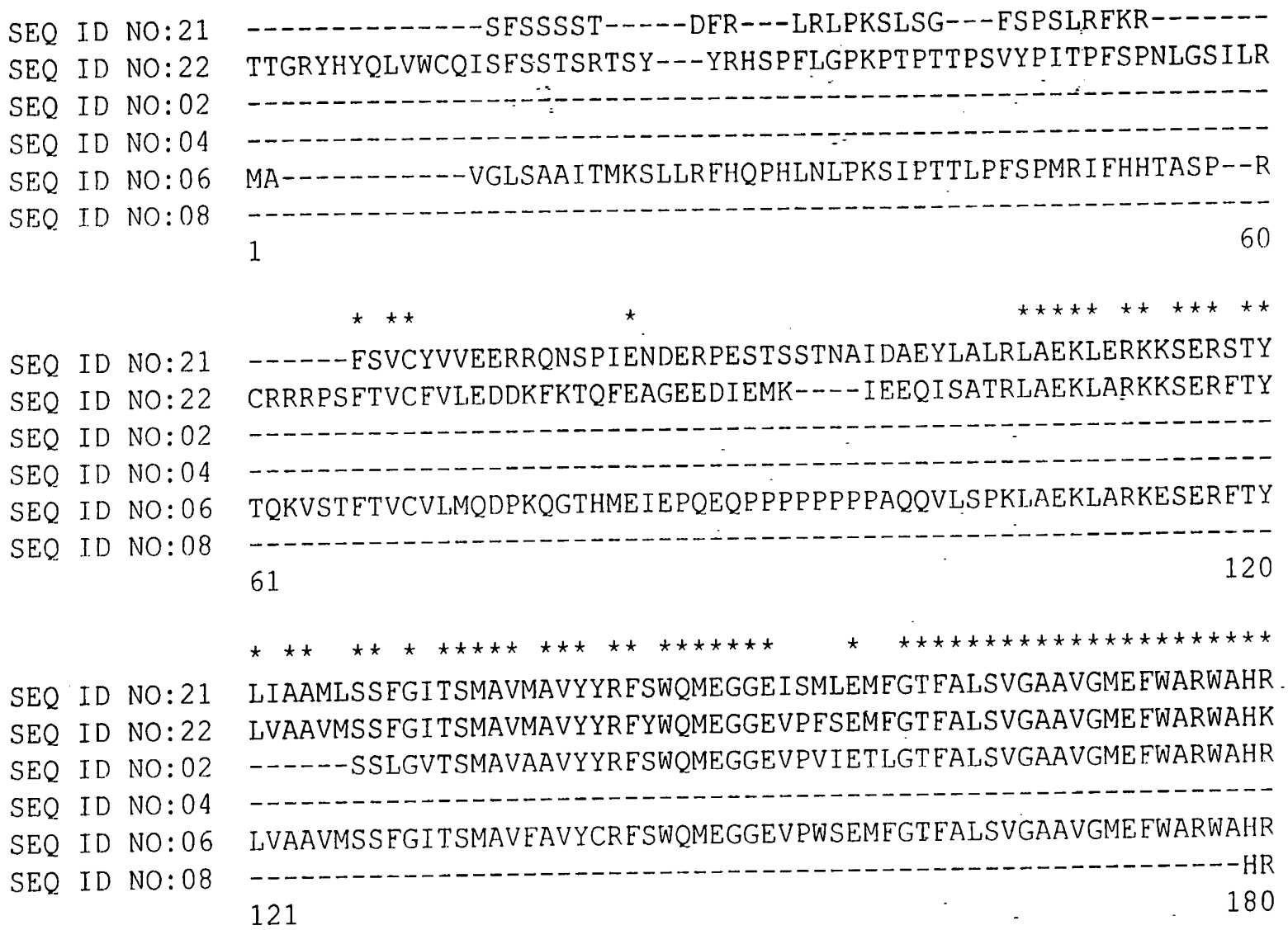

SEQ ID NO:21 ----------SFSSSST-----DFR---LRLPKSLSG---FSPSLRFKR-------
SEQ ID NO:22 TTGRYHYQLVWCQISFSSTSRTSY---YRHSPFLGPKPTPTTPSVYPITPFSPNLGSILR
SEQ ID NO:02 ----------
SEQ ID NO:04 ----------
SEQ ID NO:06 MA----------VGLSAAITMKSLLRFHQPHLNLPKSIPTTLPFSPMRIFHHTASP--R
SEQ ID NO:08 ----------
1
60
SEQ ID NO:21 ------FSVCYVVEERRQNSPIENDERPESTSSTNAIDAEYLALRLAEKLERKKSERSTY
SEQ ID NO:22 CRRRPSFTVCFVLEDDKFKTQFEAGEEDIEMK----IEEQISATRLAEKLARKKSERFTY
SEQ ID NO:02 ----------
SEQ ID NO:04 ----------
SEQ ID NO:06 TQKVSTFTVCVLMQDPKQGTHMEIEPQEQPPPPPPPPAQQVLSPKLAEKLARKESERFTY
SEQ ID NO:08 ----------
61
120
SEQ ID NO:21 LIAAMLSSFGITSMAVMAVYYRFSWQMEGGEISMLEMFGTFALSVGAAVGMEFWARWAHR
SEQ ID NO:22 LVAAVMSSFGITSMAVMAVYYRFYWQMEGGEVPFSEMFGTFALSVGAAVGMEFWARWAHK
SEQ ID NO:02 ------SSLGVTSMAVAAVYYRFSWQMEGGEVPVIETLGTFALSVGAAVGMEFWARWAHR
SEQ ID NO:04 ----------
SEQ ID NO:06 LVAAVMSSFGITSMAVFAVYCRFSWQMEGGEVPWSEMFGTFALSVGAAVGMEFWARWAHR
SEQ ID NO:08 ----------HR
121
180

FIGURE 1 - CONTINUED

```
                ******** ****** ** ********** ** **  *  ****  *  **********
SEQ ID NO:21    ALWHASLWNMHESHHKPREGPFELNDVFAIVNAGPAIGLLSYGFFNKGLVPGLCFGAGLG
SEQ ID NO:22    ALWHASLWHMHESHHKPREGPFELNDVFAIINAVPAIALLDYGFFHKGLIPGLCFGAGLG
SEQ ID NO:02    ALWHASLWHMHESHHRPREGPFELNDVFAIVNAAPAISLLAYGFFNRGIVPGLCFGAGLG
SEQ ID NO:04    ------------------------------------------------------------
SEQ ID NO:06    ALWHASLWHMHESHHRPREGPFELNDVFAIINAVPAIALLSYGFFNKGLVPGLCFGAGLG
SEQ ID NO:08    ALWHASLWDMHESHHLPRDGPFELNDVFAIVNAVPAMALLAFGFFNRGLLPGLCFGAGLG
                181                                                     240

                ** ** ****** **** ****** * *** * **  *  **  **  ***********
SEQ ID NO:21    ITVFGIAYMFVHDGLVHKRFPVGPIADVPYLRKVAAAHQLHHTDKFNGVPYGLFLGPKEL
SEQ ID NO:22    ITVFGMAYMFVHDGLVHKRFPVGPVANVPYLRKVAAAHSLHHSEKFNGVPYGLFLGPKEL
SEQ ID NO:02    ITLFGMAYMFVHDGLVHRRFPVGPIADVPYFRRVAASHKIHHMDKFGGVPYGLFLGPKEL
SEQ ID NO:04    ----------------------FANVPYFRRVAAAHQIHHMDKFEGVPYGLFLGPKEL
SEQ ID NO:06    ITVFGMAYMFVHDGLVHKRFPVGPIANVPYLRRVASAHQLHHSEKFNGVPYGLFLGPKEI
SEQ ID NO:08    ITLFGMAYMFVHXGLVHRRFPVGPIENVPYFRRVAAAHHIHHMDKFDSVPYGLFLGPKEL
                241                                                     300

                ****
SEQ ID NO:21    EEVGGNEEL-------------DKEISRRIKSYKKASG----SGSSSSS
SEQ ID NO:22    EEVGGLEEL-------------EKEVNRRTR-YIKGS-----------
SEQ ID NO:02    EEVGGLDELVSSPVSEATDTEDAGEEKTRPVVCVVRTSVFMGQSVPNEF.
SEQ ID NO:04    EEVGGIEEL-------------EKEIKRRIK--------RKETLDAIQ.
SEQ ID NO:06    EEVGGLEEL-------------EKEISR------RARSY--KIANENN.
SEQ ID NO:08    EEVGX--DLL-----------------------------------------L
                301                                             350
```

FIGURE 2

```
SEQ ID NO:23  MDTLLKTPNKLDFFIPQFHGFERLCSNNPYPSRVRLGVKKRAIKIVSSVVSGSAALLDLV
SEQ ID NO:10  ------------------------------------------------------------
SEQ ID NO:12  MGTSF-----MLFSPPPLLKPHQVPLTTPFPLPQTHHTASRNKRVHSTSKFGN--FLDFK
SEQ ID NO:14  ------------------------------------------------------------
              1                                                          60

SEQ ID NO:23  PETKKENLDFELPLYDTSKSQVVDLAIVGGGPAGLAVAQQVSEAGLSVCSIDPSPKLIWP
SEQ ID NO:10  ------------------------------------------------------------
SEQ ID NO:12  PENKPESLDFDLPWCHPSDRNRFDVIIIGAGPAGTRLAEQVSLYGVKVCCVDPDPLSVWP
SEQ ID NO:14  ------------------------------------------------------------
              61                                                         120

                                                *    * * * ** *  *       *
SEQ ID NO:23  NNYGVWVDEFEAMDLLDCLDTTWSGAVVYVDEGVKKDLSRPYGRVNRKQLKSKMLQKCIT
SEQ ID NO:10  ------------------------------GGAXSLDRPYARVARRKLXSTMMDRCVA
SEQ ID NO:12  NNYGVWRDEFESLGLEDCLDKTWPMACVYVDDGKTKYLDRCYGRVGRRKLKERLVQGCVS
SEQ ID NO:14  ------------------------------------------------------------
              121                                                        180

              *** **  **    *    *   *  **    * *  * **    * ***    * *
SEQ ID NO:25  NGVKFHQSKVTNVVHEEANSTVVCSDGVKIQASVVLDATGFSRCLVQYDKPYNPGYQVAY
SEQ ID NO:10  NGVFFHQAKVAKAVHYNASSLLICDXGVXVPASVVLXATGFSRCLVQYDKPYNPGYQFAY
SEQ ID NO:12  NGVRFHKAKVWQVQHQEFESKVLCDDGVELKGSLVVDASGFASNFVAYDKVRHHGFQIAH
SEQ ID NO:14  ------------------------------------------------------------
              181                                                        240
```


WO 99/55837 PCT/US99/08384

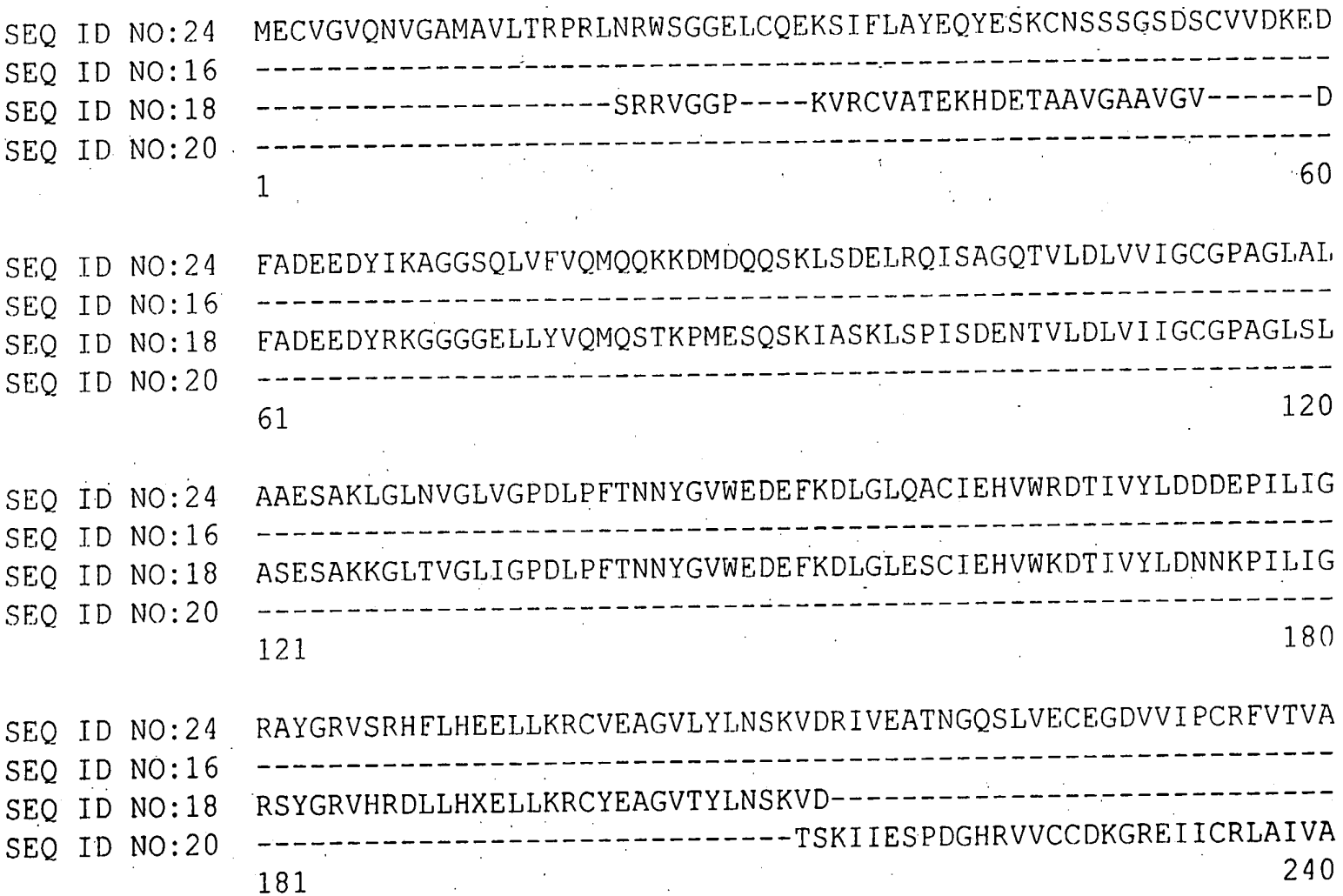
FIGURE 2 - CONTINUED

```
                 *  **   ****   *  *  ** **        *
SEQ ID NO:23  GIIAEVDGHPFDVDKMVFMDWRDKHLDSYPELKERNSKIPTFL---YAMPFSSNRIFLEE
SEQ ID NO:10  GILAEVRRHPFDIXKMLFMXWRDSHLPKGSEIRERNRRIPTFL---YAMPLLPTRIFLEE
SEQ ID NO:12  GVLAEGDDHPFDLCKMGLMGRRDSHLGNEPXLKAR---IQGFLPSSNAMPIHSNLIFLEE
SEQ ID NO:14  ----------------------------PKSVRRX-IS------------------RA
              241                                                      300

                 ***  *          ** * * ****       *  * ***** **  *  *  ****
SEQ ID NO:23  TSLVARPGLRMEDIQERMAARLKHLGINVKRIEEDERCVIPMGGPLPVLPQRVVGIGGTA
SEQ ID NO:10  TSLVARPGLAMDDIQERMAARLRHLGIRVRSVEEDDRCVIPMGGPLPVLPQRVVGIGGTA
SEQ ID NO:12  TSLVSRPVLSYMEVKRRMVARLRHLGIRVKRVLEDEKCLIPMGGPLPRIPQEVMAIGGTS
SEQ ID NO:14  TSLPARTGNSMDDIHERMAALLMHLGIRIRSVEDDERCVIPMGGSLPVLPHRVVGIGGTA
              301                                                      360

                 * **********  * ** **  *   *           *  *   **   **   *
SEQ ID NO:23  GMVHPSTGYMVARTLAAAPIVANAIVRYLGSPSSNSL---RGDQLSAEVWRDLWPIERRR
SEQ ID NO:10  GMVHPSTGYMVARTLATAPIVADAIVRFLDTGTGNGMGGLAGDALSAEVWKQLWPANRRR
SEQ ID NO:12  GVVHPSTGYMVARTMAVAPVVAFAITQCLGSTRM-----IRGKQLHDKVWNSMWPIENRL
SEQ ID NO:14  GMVHPSTGYMVARILATAPIVADSIVRFLDTANG----GIAGDALAAEVWPELWPTVTRP
              361                                                      420
```

FIGURE 2 - CONTINUED

```
                 **   ***  ******   * ** *** * * ** ******* * ***     ** ***
SEQ ID NO:23  QREFFCFGMDILLKLDLDATRRFFDAFFDLQPHYWHGFLSSRLFLPELLVFGLSLFSHAS
SEQ ID NO:10  QREFFCFGMDVLLKLDLEGTRRFFDAFFDLEPHYWHGFLSSRLFLPELLMFGLALFGNAS
SEQ ID NO:12  VREFYSFGMETLLKLDLNGSRSFFDAFFNLKPYYWQGFLSSRLTLNELLWLSISLFGHAS
SEQ ID NO:14  YRELFCFGMDVLLKLDLQGTRRFFNAFFDLEPHYWHGFLSSRLLLHELLMFGLSMFVHAS
              421                                                      480

                 * *   *  *   *  **  **
SEQ ID NO:23  NTSRLEIMTKGTVPLAKMINNLVQDR-D
SEQ ID NO:10  NSSRLEIMAKGTVPLGKMIGNLIQDRDG
SEQ ID NO:12  NPSRFDIVTKCPVPMAKMVGNIALEYIG
SEQ ID NO:14  NTSKLEIMAKGTLPLSKMVGNLIQDKDR
              481                      509
```

FIGURE 3



WO 99/55887 PCT/US99/08384

FIGURE 3 - CONTINUED

```
SEQ ID NO:24   SGAASGKFLQYELGSPRVSVQTAYGVEVEVDNNPFDPSLMVFMDYRDYLRHDAQSLEAKY
SEQ ID NO:16   ------------------------------------------------------------
SEQ ID NO:18   ------------------------------------------------------------
SEQ ID NO:20   SGAASGRLLEYEVGGPRVCVQTAYGVEVEVENNPYDPSLMVFMDYRDCFKEEFSHTEQEN
               241                                                      300

SEQ ID NO:24   PTFLYAMPMSPTRVFFEETCLASKDAMPFDLLKKKLMLRLNTLGVRIKEIYEEEWSYIPV
SEQ ID NO:16   ------------------------------------------------------------
SEQ ID NO:18   ------------------------------------------------------------
SEQ ID NO:20   PTFLYAMPMSPTRVFFEETCLASKDAMSFDLLKKRLMYRLNAMGIRILKVYEEEWSYIPV
               301                                                      360

SEQ ID NO:24   GGSLPNTEQKTLAFGAAASMVHPATGYSVVRSLSEAPKCASVLANILRQHYSKN-MLTSS
SEQ ID NO:16   ------------------------------------------------------------
SEQ ID NO:18   ------------------------------------------------------------
SEQ ID NO:20   GGSLPNTDQKNLAFGAAASMVHPATGYSVVRSLSEAPRYASVISDILGNRVPAEYMLGNS
               361                                                      420

SEQ ID NO:24   SIPSISTQAWNTLWPQERKRQRSFFLFGLALILQLDIEGIRSFFRAFFRVPKWMWQGFLG
SEQ ID NO:16   -NYSPSMLAWRTLWPXERKRQRSFFLFGLALIIQLNNEGIQTFFEAFFRVPRWMWRGFLG
SEQ ID NO:18   ------------------------------------------------------------
SEQ ID NO:20   QNYSPSMLAWRTLWPQERKRQRSFFLFGLALIIQLNNEGIQTFFEAFFRVPRWMWRGFLG
               421                                                      480
```


WO 99/55887 PCT/US99/08384

FIGURE 3 - CONTINUED

```
SEQ ID NO:24   SSLSSADLMLFAFYMFIIAPNDMRKGLIRHLLSDPTGATLIRTYLTF
SEQ ID NO:16   STLSSVDLILFSFYMFAIAPNQLRMNLVRHLLSDPTGSSMIKTYLTL
SEQ ID NO:18   ------------------------------------------------
SEQ ID NO:20   STLSSVDLILFSFYMFAIAPNQLRMNLVRHLLSDPTGSSMIKTYLTL
               481                                         528
```

SEQUENCE LISTING

<110> E. I. du Pont De Nemours and Company

<120> Carotenoid Biosynthesis Enzymes

<130> BB-1115

<140>
<141>

<150> 60/083,042
<151> April 24, 1998

<160> 24

<170> Microsoft Office 97

<210> 1
<211> 817
<212> DNA
<213> Zea mays

<400> 1
tctagcctcg gagtcacgtc catggccgtc gccgccgtgt actatcgctt cagctggcaa 60
atggagggcg gcgaggtgcc ggtgatcgag acgctgggca cgttcgcgct ctccgtcggg 120
gcggcggtcg ggatggagtt ctgggcgcgg tgggcgcacc gggcgctgtg gcacgcctcc 180
ctgtggcaca tgcacgagtc gcaccaccgg ccgcgcgagg gcccgttcga gctcaacgac 240
gtgttcgcca tcgtcaacgc cgcgccggca atctccctcc tcgcctacgg cttcttcaac 300
cgcggcatcg tgcccggcct ctgcttcggc gcgggcctgg ggattacgct gttcggcatg 360
gcctacatgt tcgtccacga cggcctggtc caccgccgct ttccggtcgg ccccatcgcc 420
gacgtgccct acttccgccg agtggctgcc tcgcacaaga tacaccacat ggacaagttc 480
ggcggcgtcc cgtatgggct cttcctggga ccaaaggagc tggaggaggt tggtggcctg 540
gacgagcttg ttagcagtcc ggtgagtgaa gctactgata ctgaagacgc aggagaagaa 600
aagacgcgtc cagttgtatg cgttgtccga acaagcgtgt tcatgggcca gagtgtgcca 660
aatgagttct agctttaggc gagtgggcca aatgagtgtg ggctgcgtcc caaggtagta 720
taagtactat gagtctgtaa taatcaggac aagaagaatt aggtaaagta atacatctat 780
cgccatcttc ctttcaaaaa aaaaaaaaaa aaaaaaa 817

<210> 2
<211> 223
<212> PRT
<213> Zea mays

<400> 2
Ser Ser Leu Gly Val Thr Ser Met Ala Val Ala Ala Val Tyr Tyr Arg
1 5 10 15

Phe Ser Trp Gln Met Glu Gly Gly Glu Val Pro Val Ile Glu Thr Leu
20 25 30

Gly Thr Phe Ala Leu Ser Val Gly Ala Ala Val Gly Met Glu Phe Trp
35 40 45

Ala Arg Trp Ala His Arg Ala Leu Trp His Ala Ser Leu Trp His Met
50 55 60

His Glu Ser His His Arg Pro Arg Glu Gly Pro Phe Glu Leu Asn Asp
65 70 75 80

```
Val Phe Ala Ile Val Asn Ala Ala Pro Ala Ile Ser Leu Leu Ala Tyr
                85                  90                  95

Gly Phe Phe Asn Arg Gly Ile Val Pro Gly Leu Cys Phe Gly Ala Gly
            100                 105                 110

Leu Gly Ile Thr Leu Phe Gly Met Ala Tyr Met Phe Val His Asp Gly
        115                 120                 125

Leu Val His Arg Arg Phe Pro Val Gly Pro Ile Ala Asp Val Pro Tyr
    130                 135                 140

Phe Arg Arg Val Ala Ala Ser His Lys Ile His His Met Asp Lys Phe
145                 150                 155                 160

Gly Gly Val Pro Tyr Gly Leu Phe Leu Gly Pro Lys Glu Leu Glu Glu
                165                 170                 175

Val Gly Gly Leu Asp Glu Leu Val Ser Ser Pro Val Ser Glu Ala Thr
            180                 185                 190

Asp Thr Glu Asp Ala Gly Glu Glu Lys Thr Arg Pro Val Val Cys Val
        195                 200                 205

Val Arg Thr Ser Val Phe Met Gly Gln Ser Val Pro Asn Glu Phe
    210                 215                 220
```

<210> 3
<211> 482
<212> DNA
<213> Oryza sativa

<220>
<221> unsure
<222> (15)

<400> 3

```
ttcgccaacg tgccntactt ccggcgggtc gccgccgccc accagataca tcacatggac  60
aagttcgaag gtgtgccata tgggctgttc cttggtccca aggagctgga ggaggtgggt 120
gggattgagg agctggagaa ggagatcaag aggaggatta agaggaaaga gaccttagat 180
gcgatccaat gagagctact tatcacacat ttttttcttt ttatttgctg tcatctctta 240
ttagtttttg cttcagtggc aattgtagcc ttgtaatatg cttttctgtc ctgttaatag 300
aagaataatt tagatagata tttggaagct catttggacg gtggagcaca caggtattag 360
aaatgatgtg ctggggctga atggctgatg cttttgctag tactgctggt aatgatatat 420
attaatcagt accattctga atttattaaa aaaaaaaaaa aaaaaaaaaa actcgatgta 480
ag                                                                  482
```

<210> 4
<211> 63
<212> PRT
<213> Oryza sativa

<400> 4

```
Phe Ala Asn Val Pro Tyr Phe Arg Arg Val Ala Ala Ala His Gln Ile
  1               5                  10                  15

His His Met Asp Lys Phe Glu Gly Val Pro Tyr Gly Leu Phe Leu Gly
            20                  25                  30

Pro Lys Glu Leu Glu Glu Val Gly Gly Ile Glu Glu Leu Glu Lys Glu
        35                  40                  45
```

Ile Lys Arg Arg Ile Lys Arg Lys Glu Thr Leu Asp Ala Ile Gln
50 55 60

<210> 5
<211> 582
<212> DNA
<213> Oryza sativa

<220>
<221> unsure
<222> (428)

<220>
<221> unsure
<222> (569)..(570)

<220>
<221> unsure
<222> (580)

<400> 5
tacagcacaa gatacaccac acggacaagt tcgagggcgt accgtatggg ctgtttcttg 60
gacccaaggt gcgtgactgg tgcctatgca tttgtcgtct atatgtgctt gtgaccagta 120
catgacaaag caatcctggt ttccttgtac cgtcttctcg atctcaacgt gtgtgtggtc 180
gtgcgtgcag gagctggagg aggttggtgg tctggaagag ctggagaagg agcttgcgag 240
aatcaaccgg agcttgtgat tttgacggcg gcaattgcag ctagcaagac tgcgcctgtc 300
gagctagcag aattgccttt gccttgcaag atttttgta tttctggccg cttcgtcgac 360
tcattaattg tcggtaccgg agagaagggg acccagaagc ggaacgcttc gagtatacac 420
atgtgaangg tgaagacaca gtaattgttt attgtggctg tggatcatcg tcctgcgctg 480
aaatttaaga gatgtcgtgg atggtatctt aattcacgcg caacttttg gtgaacttgt 540
ctcatcccgt ctgtagacac aatacgttnn aacatgtaan tt 582

<210> 6
<211> 48
<212> PRT
<213> Oryza sativa

<400> 6
Ala Arg Val Gln His Lys Ile His His Thr Asp Lys Phe Glu Gly Val
1 5 10 15

Pro Tyr Gly Leu Phe Leu Gly Pro Lys Val Glu Leu Glu Glu Val Gly
20 25 30

Gly Leu Glu Glu Leu Glu Lys Glu Leu Ala Arg Ile Asn Arg Ser Leu
35 40 45

<210> 7
<211> 604
<212> DNA
<213> Glycine max

<220>
<221> unsure
<222> (453)

<220>
<221> unsure
<222> (456)

<220>
<221> unsure
<222> (492)

<220>
<221> unsure
<222> (494)

<220>
<221> unsure
<222> (504)

<220>
<221> unsure
<222> (522)

<220>
<221> unsure
<222> (527)

<220>
<221> unsure
<222> (565)

<220>
<221> unsure
<222> (589)

<220>
<221> unsure
<222> (599)

<220>
<221> unsure
<222> (604)

<400> 7
ctctctctct ctctcatctc tctctctctc tctctctaag acacacaaac actacacatc 60
acaatctcac attgttctca agtttcaaca ctttctagtt ctaccatggc ggtaggactc 120
tccgccgcca taaccatgaa gtccctcctc cgtttccacc aacctcacct gaacctcccc 180
aaatcaatcc caacaacact ccctttctcc cccatgagaa ttttccacca cacagcatca 240
ccaagaaccc aaaaagttca actttcaccg tttgtgtcct catgcaggac ccaaaacaag 300
gcacccacat ggaaattgag ccacaagaac aacctcctcc tcctcctcct cctccccgct 360
caaacaattc tgtcaacaaa agttgggtga gaaattgggc aagaaaagat ccgaagaggt 420
cacttacctt gttgcctgct gtcaagtcta gcntcnggat cacatccaag gcaatgtcgc 480
tgttaatgca anantctcgt gggnaaatgg agggtggaga antgccntgg tctgaatgtt 540
tggcacattt gctcaacagt tgganctgcg gtgggtatgg aatttggcna atgggccana 600
ggcn 604

<210> 8
<211> 314
<212> PRT
<213> Glycine max

<400> 8
Met Ala Val Gly Leu Ser Ala Ala Ile Thr Met Lys Ser Leu Leu Arg
1 5 10 15

Phe His Gln Pro His Leu Asn Leu Pro Lys Ser Ile Pro Thr Thr Leu
20 25 30

```
Pro Phe Ser Pro Met Arg Ile Phe His His Thr Ala Ser Pro Arg Thr
        35                  40                  45

Gln Lys Val Ser Thr Phe Thr Val Cys Val Leu Met Gln Asp Pro Lys
    50                  55                  60

Gln Gly Thr His Met Glu Ile Glu Pro Gln Glu Gln Pro Pro Pro Pro
65                  70                  75                  80

Pro Pro Pro Pro Ala Gln Gln Val Leu Ser Pro Lys Leu Ala Glu Lys
                85                  90                  95

Leu Ala Arg Lys Glu Ser Glu Arg Phe Thr Tyr Leu Val Ala Ala Val
            100                 105                 110

Met Ser Ser Phe Gly Ile Thr Ser Met Ala Val Phe Ala Val Tyr Cys
        115                 120                 125

Arg Phe Ser Trp Gln Met Glu Gly Gly Glu Val Pro Trp Ser Glu Met
    130                 135                 140

Phe Gly Thr Phe Ala Leu Ser Val Gly Ala Ala Val Gly Met Glu Phe
145                 150                 155                 160

Trp Ala Arg Trp Ala His Arg Ala Leu Trp His Ala Ser Leu Trp His
                165                 170                 175

Met His Glu Ser His His Arg Pro Arg Glu Gly Pro Phe Glu Leu Asn
            180                 185                 190

Asp Val Phe Ala Ile Ile Asn Ala Val Pro Ala Ile Ala Leu Leu Ser
        195                 200                 205

Tyr Gly Phe Phe Asn Lys Gly Leu Val Pro Gly Leu Cys Phe Gly Ala
    210                 215                 220

Gly Leu Gly Ile Thr Val Phe Gly Met Ala Tyr Met Phe Val His Asp
225                 230                 235                 240

Gly Leu Val His Lys Arg Phe Pro Val Gly Pro Ile Ala Asn Val Pro
                245                 250                 255

Tyr Leu Arg Arg Val Ala Ser Ala His Gln Leu His His Ser Glu Lys
            260                 265                 270

Phe Asn Gly Val Pro Tyr Gly Leu Phe Leu Gly Pro Lys Glu Ile Glu
        275                 280                 285

Glu Val Gly Gly Leu Glu Glu Leu Glu Lys Glu Ile Ser Arg Arg Ala
    290                 295                 300

Arg Ser Tyr Lys Ile Ala Asn Glu Asn Asn
305                 310
```

<210> 10
<211> 314
<212> PRT
<213> Glycine max

<400> 10

```
Met Ala Val Gly Leu Ser Ala Ala Ile Thr Met Lys Ser Leu Leu Arg
 1               5                  10                  15

Phe His Gln Pro His Leu Asn Leu Pro Lys Ser Ile Pro Thr Thr Leu
            20                  25                  30

Pro Phe Ser Pro Met Arg Ile Phe His His Thr Ala Ser Pro Arg Thr
        35                  40                  45

Gln Lys Val Ser Thr Phe Thr Val Cys Val Leu Met Gln Asp Pro Lys
    50                  55                  60

Gln Gly Thr His Met Glu Ile Glu Pro Gln Glu Gln Pro Pro Pro Pro
65                  70                  75                  80

Pro Pro Pro Pro Ala Gln Gln Val Leu Ser Pro Lys Leu Ala Glu Lys
                85                  90                  95

Leu Ala Arg Lys Glu Ser Glu Arg Phe Thr Tyr Leu Val Ala Ala Val
            100                 105                 110

Met Ser Ser Phe Gly Ile Thr Ser Met Ala Val Phe Ala Val Tyr Cys
        115                 120                 125

Arg Phe Ser Trp Gln Met Glu Gly Gly Glu Val Pro Trp Ser Glu Met
    130                 135                 140

Phe Gly Thr Phe Ala Leu Ser Val Gly Ala Ala Val Gly Met Glu Phe
145                 150                 155                 160

Trp Ala Arg Trp Ala His Arg Ala Leu Trp His Ala Ser Leu Trp His
                165                 170                 175

Met His Glu Ser His His Arg Pro Arg Glu Gly Pro Phe Glu Leu Asn
            180                 185                 190

Asp Val Phe Ala Ile Ile Asn Ala Val Pro Ala Ile Ala Leu Leu Ser
        195                 200                 205

Tyr Gly Phe Phe Asn Lys Gly Leu Val Pro Gly Leu Cys Phe Gly Ala
    210                 215                 220

Gly Leu Gly Ile Thr Val Phe Gly Met Ala Tyr Met Phe Val His Asp
225                 230                 235                 240

Gly Leu Val His Lys Arg Phe Pro Val Gly Pro Ile Ala Asn Val Pro
                245                 250                 255

Tyr Leu Arg Arg Val Ala Ser Ala His Gln Leu His His Ser Glu Lys
            260                 265                 270

Phe Asn Gly Val Pro Tyr Gly Leu Phe Leu Gly Pro Lys Glu Ile Glu
        275                 280                 285

Glu Val Gly Gly Leu Glu Glu Leu Glu Lys Glu Ile Ser Arg Arg Ala
    290                 295                 300

Arg Ser Tyr Lys Ile Ala Asn Glu Asn Asn
305                 310
```

<210> 12
<211> 254
<212> PRT
<213> Glycine max

<400> 12

```
Met Gln Asp Pro Lys Gln Gly Thr His Met Glu Ile Glu Pro Gln Glu
 1               5                  10                  15

Gln Pro Pro Pro Pro Pro Pro Pro Ala Gln Gln Val Leu Ser Pro Lys
            20                  25                  30

Leu Ala Glu Lys Leu Ala Arg Lys Glu Ser Glu Arg Phe Thr Tyr Leu
        35                  40                  45

Val Ala Ala Val Met Ser Ser Phe Gly Ile Thr Ser Met Ala Val Phe
    50                  55                  60

Ala Val Tyr Cys Arg Phe Ser Trp Gln Met Glu Gly Gly Glu Val Pro
65                  70                  75                  80

Trp Ser Glu Met Phe Gly Thr Phe Ala Leu Ser Val Gly Ala Ala Val
                85                  90                  95

Gly Met Glu Phe Trp Ala Arg Trp Ala His Arg Ala Leu Trp His Ala
            100                 105                 110

Ser Leu Trp His Met His Glu Ser His His Arg Pro Arg Glu Gly Pro
        115                 120                 125

Phe Glu Leu Asn Asp Val Phe Ala Ile Ile Asn Ala Val Pro Ala Ile
    130                 135                 140

Ala Leu Leu Ser Tyr Gly Phe Phe Asn Lys Gly Leu Val Pro Gly Leu
145                 150                 155                 160

Cys Phe Gly Ala Gly Leu Gly Ile Thr Val Phe Gly Met Ala Tyr Met
                165                 170                 175

Phe Val His Asp Gly Leu Val His Lys Arg Phe Pro Val Gly Pro Ile
            180                 185                 190

Ala Asn Val Pro Tyr Leu Arg Arg Val Ala Ser Ala His Gln Leu His
        195                 200                 205

His Ser Glu Lys Phe Asn Gly Val Pro Tyr Gly Leu Phe Leu Gly Pro
    210                 215                 220

Lys Glu Ile Glu Glu Val Gly Gly Leu Glu Glu Leu Glu Lys Glu Ile
225                 230                 235                 240

Ser Arg Arg Ala Arg Ser Tyr Lys Ile Ala Asn Glu Asn Asn
                245                 250
```

<210> 13
<211> 396
<212> DNA
<213> Triticum aestivum

<220>
<221> unsure
<222> (224)

<220>
<221> unsure
<222> (360)

<400> 13
gcaccgcgcg ctgtggcacg cgtcgctgtg ggacatgcac gagtcgcacc acctcccccg 60
cgacgggccc ttcgagctca acgacgtgtt cgccatcgtc aacgccgtcc ccgccatggc 120
cctcctcgcc ttcggcttct tcaaccgcgg cctcctcccc ggcctctgct tcggcgccgg 180
tctggggatc acgctgttcg ggatggccta catgttcgtc cacnacggcc tggtgcaccg 240
ccgcttcccc gtgggcccca tcgagaacgt gccctacttc cggcgagtcg ccgccgcgca 300
ccatattcat cacatggaca agttcgacag cgtgccatac gggctcttcc tcggccccaa 360
ggagctggag gaggtggggn aggatcttct tctttt 396

<210> 14
<211> 131
<212> PRT
<213> Triticum aestivum

<220>
<221> UNSURE
<222> (75)

<220>
<221> UNSURE
<222> (127)

<400> 14
His Arg Ala Leu Trp His Ala Ser Leu Trp Asp Met His Glu Ser His
1 5 10 15

His Leu Pro Arg Asp Gly Pro Phe Glu Leu Asn Asp Val Phe Ala Ile
20 25 30

Val Asn Ala Val Pro Ala Met Ala Leu Leu Ala Phe Gly Phe Phe Asn
35 40 45

Arg Gly Leu Leu Pro Gly Leu Cys Phe Gly Ala Gly Leu Gly Ile Thr
50 55 60

Leu Phe Gly Met Ala Tyr Met Phe Val His Xaa Gly Leu Val His Arg
65 70 75 80

Arg Phe Pro Val Gly Pro Ile Glu Asn Val Pro Tyr Phe Arg Arg Val
85 90 95

Ala Ala Ala His His Ile His His Met Asp Lys Phe Asp Ser Val Pro
100 105 110

Tyr Gly Leu Phe Leu Gly Pro Lys Glu Leu Glu Glu Val Gly Xaa Asp
115 120 125

Leu Leu Leu
130

<210> 15
<211> 1358

<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (12)

<220>
<221> unsure
<222> (57)

<220>
<221> unsure
<222> (160)

<220>
<221> unsure
<222> (169)

<220>
<221> unsure
<222> (193)

<220>
<221> unsure
<222> (305)

<220>
<221> unsure
<222> (324)

<400> 15
ggcggcgcca antcgctcga ccgcccctac gcccgcgtcg cgcgccgcaa gctcaantcc 60
accatgatgg accgctgcgt cgccaacggc gtcttcttcc accaggccaa ggtcgccaag 120
gccgtccact acaacgcctc gtcgctcctc atctgtgacn acggcgtcnc cgtcccggca 180
agcgtcgtgc tcnacgccac gggcttctcg cgctgcctcg tgcagtacga caagccgtac 240
aacccgggtt accagttcgc ctacggcatc ctcgccgaag ttcgacgcca cccgttcgac 300
attcnaaaaa tgctcttcat ggantggcgc gactcgcacc tccccaaagg gtcggaaatc 360
agggagcgca accgccgcat ccccaccttc ctctacgcca tgccccttct ccccaccagg 420
atcttcctcg aggagacgtc cctggtcgcg cgcccggggc tcgccatgga cgacatccag 480
gagcgcatgg ccgcgcgcct caggcacctg ggtatacgcg tccgaagcgt cgaggaggac 540
gatcgctgcg tcatccccat gggagggccg ctgcccgtcc tgccgcagag ggttgtcggc 600
atcggcggca cggcagggat ggtgcacccg tccacgggct acatggtggc gcgcacgctt 660
gccaccgcgc ctatcgtggc ggacgccatc gtaaggttcc tcgacaccgg caccggcaac 720
ggcatgggtg gcctggcagg ggacgcgctc tccgccgagg tgtggaagca gctgtggcca 780
gccaacaggc ggcggcagag ggagttcttc tgcttcggca tggacgtcct gctcaagctg 840
gacctcgagg gaacgcggcg gttcttcgac gccttcttcg acctggagcc acactactgg 900
cacggtttcc tgtcatccag actgttcctg ccggagctct tgatgttcgg cctcgcactg 960
ttcgggaacg cctccaactc gtcgaggctg gagatcatgg ccaagggcac cgtgcctctt 1020
ggcaagatga ttggcaactt gatacaggac agagatgggt gaggagggta tgtataccta 1080
catttttcac gtgaagatct gatctccatt ggatctctga ttttggtatc gatgattttc 1140
actgtattta cgatttgcaa agatggattc acaaaacaca gttagcaaca gcagttcagg 1200
acctcctgtc agatatagga attgctgctg caacgctact tcagtatggt gattacagag 1260
gtgtatagtt gcacttgcac actgagggat gtcgtgagaa tctacgtatc agatatcatg 1320
gtcttcatat aaaagatcaa atttccaaca aaaaaaaa 1358

<210> 16
<211> 353
<212> PRT
<213> Zea mays

<220>
<221> UNSURE
<222> (4)

<220>
<221> UNSURE
<222> (19)

<220>
<221> UNSURE
<222> (54)

<220>
<221> UNSURE
<222> (57)

<220>
<221> UNSURE
<222> (65)

<220>
<221> UNSURE
<222> (102)

<220>
<221> UNSURE
<222> (108)

<400> 16

```
Gly Gly Ala Xaa Ser Leu Asp Arg Pro Tyr Ala Arg Val Ala Arg Arg
 1               5                  10                  15

Lys Leu Xaa Ser Thr Met Met Asp Arg Cys Val Ala Asn Gly Val Phe
            20                  25                  30

Phe His Gln Ala Lys Val Ala Lys Ala Val His Tyr Asn Ala Ser Ser
        35                  40                  45

Leu Leu Ile Cys Asp Xaa Gly Val Xaa Val Pro Ala Ser Val Val Leu
    50                  55                  60

Xaa Ala Thr Gly Phe Ser Arg Cys Leu Val Gln Tyr Asp Lys Pro Tyr
65                  70                  75                  80

Asn Pro Gly Tyr Gln Phe Ala Tyr Gly Ile Leu Ala Glu Val Arg Arg
                85                  90                  95

His Pro Phe Asp Ile Xaa Lys Met Leu Phe Met Xaa Trp Arg Asp Ser
            100                 105                 110

His Leu Pro Lys Gly Ser Glu Ile Arg Glu Arg Asn Arg Arg Ile Pro
        115                 120                 125

Thr Phe Leu Tyr Ala Met Pro Leu Leu Pro Thr Arg Ile Phe Leu Glu
    130                 135                 140

Glu Thr Ser Leu Val Ala Arg Pro Gly Leu Ala Met Asp Asp Ile Gln
145                 150                 155                 160

Glu Arg Met Ala Ala Arg Leu Arg His Leu Gly Ile Arg Val Arg Ser
                165                 170                 175
```

```
Val Glu Glu Asp Asp Arg Cys Val Ile Pro Met Gly Gly Pro Leu Pro
            180                 185                 190

Val Leu Pro Gln Arg Val Val Gly Ile Gly Gly Thr Ala Gly Met Val
        195                 200                 205

His Pro Ser Thr Gly Tyr Met Val Ala Arg Thr Leu Ala Thr Ala Pro
    210                 215                 220

Ile Val Ala Asp Ala Ile Val Arg Phe Leu Asp Thr Gly Thr Gly Asn
225                 230                 235                 240

Gly Met Gly Gly Leu Ala Gly Asp Ala Leu Ser Ala Glu Val Trp Lys
                245                 250                 255

Gln Leu Trp Pro Ala Asn Arg Arg Arg Gln Arg Glu Phe Phe Cys Phe
            260                 265                 270

Gly Met Asp Val Leu Leu Lys Leu Asp Leu Glu Gly Thr Arg Arg Phe
        275                 280                 285

Phe Asp Ala Phe Phe Asp Leu Glu Pro His Tyr Trp His Gly Phe Leu
    290                 295                 300

Ser Ser Arg Leu Phe Leu Pro Glu Leu Leu Met Phe Gly Leu Ala Leu
305                 310                 315                 320

Phe Gly Asn Ala Ser Asn Ser Ser Arg Leu Glu Ile Met Ala Lys Gly
                325                 330                 335

Thr Val Pro Leu Gly Lys Met Ile Gly Asn Leu Ile Gln Asp Arg Asp
            340                 345                 350

Gly
```

<210> 17
<211> 1731
<212> DNA
<213> Glycine max

<220>
<221> unsure
<222> (878)

<400> 17

```
gcacgagctt ccaattacca aagtctcttc actatagaca acacatttct gcacaatttt   60
attgctactt ttctccatca cctaaaaaat gggcactagt ttcatgctat tttctccacc  120
gcctctgctc aagcctcacc aagtacccct caccactcca ttccctcttc ctcaaaccca  180
tcacacagca tcaagaaaca agagggtcca cagcaccagc aaattcggaa actttcttga  240
cttcaaaccc gagaacaaac ccgaatcctt agattttgac cttccatggt gccacccttc  300
tgaccgcaat cgttttgatg tgatcatcat tggtgctggc cctgcaggca caaggctcgc  360
cgagcaagtg tccctctatg gagttaaggt ttgttgtgtt gatcctgacc ctctttctgt  420
gtggcctaac aactatggtg tgtggagaga tgagtttgag agccttggtc ttgaggattg  480
cttggacaaa acatggccca tggcttgtgt ttatgtggat gatggcaaga ccaagtattt  540
ggatcggtgt tatgggaggg ttggtaggag gaagctgaag gagagattgg ttcaaggctg  600
tgtctccaat ggggttaggt ttcacaaggc aaaggtgtgg caagttcagc accaagagtt  660
tgagtccaaa gttttgtgtg atgatggggt ggaattgaag gggagtttgg ttgttgatgc  720
tagtggattc gcaagtaatt ttgtagcata tgacaaggtg agacaccatg gttttcagat  780
tgctcatggt gttttggctg aaggggatga tcacccttt gatttgtgca aaatgggttt  840
```

```
aatgggacgg agggactctc atttggggaa tgaacctnac ttgaaagcta gaattcaagg  900
gttcctacct tcctctaatg caatgccaat tcattccaat ttgatatttc ttgaggagac  960
ttcccttgtg agccgtccag tgttgtctta catggaagtg aaaaggagga tggttgcaag 1020
gctaaggcac ttaggcatta gagtgaagag ggttttggag gatgaaaagt gtttgattcc 1080
aatgggagga cctcttccaa ggatccccca agaagtgatg gctattggtg gcacttctgg 1140
ggtagtgcac ccttccacag gttacatggt ggctagaaca atggctgtgg ctcctgttgt 1200
ggcttttgct ataactcaat gccttggctc caccagaatg atcaggggga aacaacttca 1260
tgataaagtt tggaatagca tgtggccaat tgagaacagg ctggtaaggg aattttactc 1320
ttttggaatg gagactttgt tgaaacttga cttgaatgga agcaggagtt tctttgatgc 1380
atttttaac ttgaaacctt attactggca agggttcctc tcttcgaggt tgactttgaa 1440
tgagcttctt tggttaagca tatcactctt tggacatgcc tcaaatccat ccaggtttga 1500
tattgtcaca aagtgtcctg ttccaatggc taagatggtg ggcaatatcg ctttggagta 1560
cattggatga tgatgatgat gatgatggaa ggttgtgaaa tgcgtgttta agtttttgtt 1620
tttcattggc tgaacttgca tcttgatttg ttgagttgtc attgaattga tatatatgat 1680
aatttagctg ctaataaaaa attaattcag ttccgtttaa aaaaaaaaaa a           1731

<210>  18
<211>  493
<212>  PRT
<213>  Glycine max

<220>
<221>  UNSURE
<222>  (264)

<400>  18
Met Gly Thr Ser Phe Met Leu Phe Ser Pro Pro Pro Leu Leu Lys Pro
  1               5                  10                  15

His Gln Val Pro Leu Thr Thr Pro Phe Pro Leu Pro Gln Thr His His
             20                  25                  30

Thr Ala Ser Arg Asn Lys Arg Val His Ser Thr Ser Lys Phe Gly Asn
         35                  40                  45

Phe Leu Asp Phe Lys Pro Glu Asn Lys Pro Glu Ser Leu Asp Phe Asp
     50                  55                  60

Leu Pro Trp Cys His Pro Ser Asp Arg Asn Arg Phe Asp Val Ile Ile
 65                  70                  75                  80

Ile Gly Ala Gly Pro Ala Gly Thr Arg Leu Ala Glu Gln Val Ser Leu
                 85                  90                  95

Tyr Gly Val Lys Val Cys Cys Val Asp Pro Asp Pro Leu Ser Val Trp
            100                 105                 110

Pro Asn Asn Tyr Gly Val Trp Arg Asp Glu Phe Glu Ser Leu Gly Leu
        115                 120                 125

Glu Asp Cys Leu Asp Lys Thr Trp Pro Met Ala Cys Val Tyr Val Asp
    130                 135                 140

Asp Gly Lys Thr Lys Tyr Leu Asp Arg Cys Tyr Gly Arg Val Gly Arg
145                 150                 155                 160

Arg Lys Leu Lys Glu Arg Leu Val Gln Gly Cys Val Ser Asn Gly Val
                165                 170                 175

Arg Phe His Lys Ala Lys Val Trp Gln Val Gln His Gln Glu Phe Glu
            180                 185                 190
```

Ser Lys Val Leu Cys Asp Asp Gly Val Glu Leu Lys Gly Ser Leu Val
195 200 205

Val Asp Ala Ser Gly Phe Ala Ser Asn Phe Val Ala Tyr Asp Lys Val
210 215 220

Arg His His Gly Phe Gln Ile Ala His Gly Val Leu Ala Glu Gly Asp
225 230 235 240

Asp His Pro Phe Asp Leu Cys Lys Met Gly Leu Met Gly Arg Arg Asp
245 250 255

Ser His Leu Gly Asn Glu Pro Xaa Leu Lys Ala Arg Ile Gln Gly Phe
260 265 270

Leu Pro Ser Ser Asn Ala Met Pro Ile His Ser Asn Leu Ile Phe Leu
275 280 285

Glu Glu Thr Ser Leu Val Ser Arg Pro Val Leu Ser Tyr Met Glu Val
290 295 300

Lys Arg Arg Met Val Ala Arg Leu Arg His Leu Gly Ile Arg Val Lys
305 310 315 320

Arg Val Leu Glu Asp Glu Lys Cys Leu Ile Pro Met Gly Gly Pro Leu
325 330 335

Pro Arg Ile Pro Gln Glu Val Met Ala Ile Gly Gly Thr Ser Gly Val
340 345 350

Val His Pro Ser Thr Gly Tyr Met Val Ala Arg Thr Met Ala Val Ala
355 360 365

Pro Val Val Ala Phe Ala Ile Thr Gln Cys Leu Gly Ser Thr Arg Met
370 375 380

Ile Arg Gly Lys Gln Leu His Asp Lys Val Trp Asn Ser Met Trp Pro
385 390 395 400

Ile Glu Asn Arg Leu Val Arg Glu Phe Tyr Ser Phe Gly Met Glu Thr
405 410 415

Leu Leu Lys Leu Asp Leu Asn Gly Ser Arg Ser Phe Phe Asp Ala Phe
420 425 430

Phe Asn Leu Lys Pro Tyr Tyr Trp Gln Gly Phe Leu Ser Ser Arg Leu
435 440 445

Thr Leu Asn Glu Leu Leu Trp Leu Ser Ile Ser Leu Phe Gly His Ala
450 455 460

Ser Asn Pro Ser Arg Phe Asp Ile Val Thr Lys Cys Pro Val Pro Met
465 470 475 480

Ala Lys Met Val Gly Asn Ile Ala Leu Glu Tyr Ile Gly
485 490

<210> 19
<211> 853

<212> DNA
<213> Triticum aestivum

<400> 19
taccgaagtc ggtacggcgt nccatctcgc gagcgacgtc gctgcccgcg cgcacgggga 60
actccatgga cgacatccac gagcgcatgg ccgcgctcct gatgcacctg ggaattcgca 120
tccggagcgt cgaggacgac gagcgctgcg tgatccccat gggcgggtcg ctgcccgtgc 180
tgccgcacag ggtggtgggc atcggcggca cggccgggat ggtgcatccg tccacagggt 240
acatggtggc gcgcatattg gcgaccgcgc ccatcgtggc agactccatt gtgcggtttc 300
tagacactgc aaatggtggc atcgcccggg acgcgctcgc cgccgaggtg tggcccgagc 360
tgtggccgac ggttacgcgg ccgtacaggg aactcttctg cttcggcatg gacgtcctgc 420
tcaagctgga cctccaaggt acacgacgct tcttcaacgc attcttcgac cttgagccgc 480
actattggca cggcttcctc tcgtcgaggc tgctcctgca tgagctcctg atgttcgggc 540
tctcgatgtt cgtgcacgct tccaacacgt ccaagctgga gatcatggcc aagggcaccc 600
tgcctctttc caagatggtc ggcaacttga tacaggacaa ggataggtga tgacttagag 660
ggtatgtatg tacctgcatc tcaagatctt catggggtct ttgattttgg catagctttt 720
tctcaaatgt atctatgatt ggcaaagagg atttaaatag cagtggtagc aacagcagct 780
taggacctcc agagagatgt aacaattctt gctgttgcta cgttattgag atgtatattt 840
cataaaaaaa aaa 853

<210> 20
<211> 215
<212> PRT
<213> Triticum aestivum

<220>
<221> UNSURE
<222> (7)

<400> 20
Pro Lys Ser Val Arg Arg Xaa Ile Ser Arg Ala Thr Ser Leu Pro Ala
1 5 10 15

Arg Thr Gly Asn Ser Met Asp Asp Ile His Glu Arg Met Ala Ala Leu
20 25 30

Leu Met His Leu Gly Ile Arg Ile Arg Ser Val Glu Asp Asp Glu Arg
35 40 45

Cys Val Ile Pro Met Gly Gly Ser Leu Pro Val Leu Pro His Arg Val
50 55 60

Val Gly Ile Gly Gly Thr Ala Gly Met Val His Pro Ser Thr Gly Tyr
65 70 75 80

Met Val Ala Arg Ile Leu Ala Thr Ala Pro Ile Val Ala Asp Ser Ile
85 90 95

Val Arg Phe Leu Asp Thr Ala Asn Gly Gly Ile Ala Gly Asp Ala Leu
100 105 110

Ala Ala Glu Val Trp Pro Glu Leu Trp Pro Thr Val Thr Arg Pro Tyr
115 120 125

Arg Glu Leu Phe Cys Phe Gly Met Asp Val Leu Leu Lys Leu Asp Leu
130 135 140

Gln Gly Thr Arg Arg Phe Phe Asn Ala Phe Phe Asp Leu Glu Pro His
145 150 155 160

Tyr Trp His Gly Phe Leu Ser Ser Arg Leu Leu Leu His Glu Leu Leu
165 170 175

Met Phe Gly Leu Ser Met Phe Val His Ala Ser Asn Thr Ser Lys Leu
180 185 190

Glu Ile Met Ala Lys Gly Thr Leu Pro Leu Ser Lys Met Val Gly Asn
195 200 205

Leu Ile Gln Asp Lys Asp Arg
210 215

<210> 21
<211> 294
<212> PRT
<213> Arabidopsis thaliana

<400> 21
Ser Phe Ser Ser Ser Ser Thr Asp Phe Arg Leu Arg Leu Pro Lys Ser
1 5 10 15

Leu Ser Gly Phe Ser Pro Ser Leu Arg Phe Lys Arg Phe Ser Val Cys
20 25 30

Tyr Val Val Glu Glu Arg Arg Gln Asn Ser Pro Ile Glu Asn Asp Glu
35 40 45

Arg Pro Glu Ser Thr Ser Ser Thr Asn Ala Ile Asp Ala Glu Tyr Leu
50 55 60

Ala Leu Arg Leu Ala Glu Lys Leu Glu Arg Lys Lys Ser Glu Arg Ser
65 70 75 80

Thr Tyr Leu Ile Ala Ala Met Leu Ser Ser Phe Gly Ile Thr Ser Met
85 90 95

Ala Val Met Ala Val Tyr Tyr Arg Phe Ser Trp Gln Met Glu Gly Gly
100 105 110

Glu Ile Ser Met Leu Glu Met Phe Gly Thr Phe Ala Leu Ser Val Gly
115 120 125

Ala Ala Val Gly Met Glu Phe Trp Ala Arg Trp Ala His Arg Ala Leu
130 135 140

Trp His Ala Ser Leu Trp Asn Met His Glu Ser His His Lys Pro Arg
145 150 155 160

Glu Gly Pro Phe Glu Leu Asn Asp Val Phe Ala Ile Val Asn Ala Gly
165 170 175

Pro Ala Ile Gly Leu Leu Ser Tyr Gly Phe Phe Asn Lys Gly Leu Val
180 185 190

Pro Gly Leu Cys Phe Gly Ala Gly Leu Gly Ile Thr Val Phe Gly Ile
195 200 205

Ala Tyr Met Phe Val His Asp Gly Leu Val His Lys Arg Phe Pro Val
210 215 220

Gly Pro Ile Ala Asp Val Pro Tyr Leu Arg Lys Val Ala Ala Ala His
225 230 235 240

Gln Leu His His Thr Asp Lys Phe Asn Gly Val Pro Tyr Gly Leu Phe
245 250 255

Leu Gly Pro Lys Glu Leu Glu Glu Val Gly Gly Asn Glu Glu Leu Asp
260 265 270

Lys Glu Ile Ser Arg Arg Ile Lys Ser Tyr Lys Lys Ala Ser Gly Ser
275 280 285

Gly Ser Ser Ser Ser Ser
290

<210> 22
<211> 316
<212> PRT
<213> Capsicum annuum

<400> 22
Thr Thr Gly Arg Tyr His Tyr Gln Leu Val Trp Cys Gln Ile Ser Phe
1 5 10 15

Ser Ser Thr Ser Arg Thr Ser Tyr Tyr Arg His Ser Pro Phe Leu Gly
20 25 30

Pro Lys Pro Thr Pro Thr Thr Pro Ser Val Tyr Pro Ile Thr Pro Phe
35 40 45

Ser Pro Asn Leu Gly Ser Ile Leu Arg Cys Arg Arg Arg Pro Ser Phe
50 55 60

Thr Val Cys Phe Val Leu Glu Asp Asp Lys Phe Lys Thr Gln Phe Glu
65 70 75 80

Ala Gly Glu Glu Asp Ile Glu Met Lys Ile Glu Glu Gln Ile Ser Ala
85 90 95

Thr Arg Leu Ala Glu Lys Leu Ala Arg Lys Lys Ser Glu Arg Phe Thr
100 105 110

Tyr Leu Val Ala Ala Val Met Ser Ser Phe Gly Ile Thr Ser Met Ala
115 120 125

Val Met Ala Val Tyr Tyr Arg Phe Tyr Trp Gln Met Glu Gly Gly Glu
130 135 140

Val Pro Phe Ser Glu Met Phe Gly Thr Phe Ala Leu Ser Val Gly Ala
145 150 155 160

Ala Val Gly Met Glu Phe Trp Ala Arg Trp Ala His Lys Ala Leu Trp
165 170 175

His Ala Ser Leu Trp His Met His Glu Ser His His Lys Pro Arg Glu
180 185 190

Gly Pro Phe Glu Leu Asn Asp Val Phe Ala Ile Ile Asn Ala Val Pro
195 200 205

Ala Ile Ala Leu Leu Asp Tyr Gly Phe Phe His Lys Gly Leu Ile Pro
210 215 220

Gly Leu Cys Phe Gly Ala Gly Leu Gly Ile Thr Val Phe Gly Met Ala
225 230 235 240

Tyr Met Phe Val His Asp Gly Leu Val His Lys Arg Phe Pro Val Gly
245 250 255

Pro Val Ala Asn Val Pro Tyr Leu Arg Lys Val Ala Ala Ala His Ser
260 265 270

Leu His His Ser Glu Lys Phe Asn Gly Val Pro Tyr Gly Leu Phe Leu
275 280 285

Gly Pro Lys Glu Leu Glu Glu Val Gly Gly Leu Glu Glu Leu Glu Lys
290 295 300

Glu Val Asn Arg Arg Thr Arg Tyr Ile Lys Gly Ser
305 310 315

<210> 23
<211> 501
<212> PRT
<213> Arabidopsis thaliana

<400> 23
Met Asp Thr Leu Leu Lys Thr Pro Asn Lys Leu Asp Phe Phe Ile Pro
1 5 10 15

Gln Phe His Gly Phe Glu Arg Leu Cys Ser Asn Asn Pro Tyr Pro Ser
20 25 30

Arg Val Arg Leu Gly Val Lys Lys Arg Ala Ile Lys Ile Val Ser Ser
35 40 45

Val Val Ser Gly Ser Ala Ala Leu Leu Asp Leu Val Pro Glu Thr Lys
50 55 60

Lys Glu Asn Leu Asp Phe Glu Leu Pro Leu Tyr Asp Thr Ser Lys Ser
65 70 75 80

Gln Val Val Asp Leu Ala Ile Val Gly Gly Gly Pro Ala Gly Leu Ala
85 90 95

Val Ala Gln Gln Val Ser Glu Ala Gly Leu Ser Val Cys Ser Ile Asp
100 105 110

Pro Ser Pro Lys Leu Ile Trp Pro Asn Asn Tyr Gly Val Trp Val Asp
115 120 125

Glu Phe Glu Ala Met Asp Leu Leu Asp Cys Leu Asp Thr Thr Trp Ser
130 135 140

Gly Ala Val Val Tyr Val Asp Glu Gly Val Lys Lys Asp Leu Ser Arg
145 150 155 160

Pro Tyr Gly Arg Val Asn Arg Lys Gln Leu Lys Ser Lys Met Leu Gln
165 170 175

```
Lys Cys Ile Thr Asn Gly Val Lys Phe His Gln Ser Lys Val Thr Asn
            180                 185                 190

Val Val His Glu Glu Ala Asn Ser Thr Val Val Cys Ser Asp Gly Val
        195                 200                 205

Lys Ile Gln Ala Ser Val Val Leu Asp Ala Thr Gly Phe Ser Arg Cys
    210                 215                 220

Leu Val Gln Tyr Asp Lys Pro Tyr Asn Pro Gly Tyr Gln Val Ala Tyr
225                 230                 235                 240

Gly Ile Ile Ala Glu Val Asp Gly His Pro Phe Asp Val Asp Lys Met
                245                 250                 255

Val Phe Met Asp Trp Arg Asp Lys His Leu Asp Ser Tyr Pro Glu Leu
            260                 265                 270

Lys Glu Arg Asn Ser Lys Ile Pro Thr Phe Leu Tyr Ala Met Pro Phe
        275                 280                 285

Ser Ser Asn Arg Ile Phe Leu Glu Glu Thr Ser Leu Val Ala Arg Pro
    290                 295                 300

Gly Leu Arg Met Glu Asp Ile Gln Glu Arg Met Ala Ala Arg Leu Lys
305                 310                 315                 320

His Leu Gly Ile Asn Val Lys Arg Ile Glu Glu Asp Glu Arg Cys Val
                325                 330                 335

Ile Pro Met Gly Gly Pro Leu Pro Val Leu Pro Gln Arg Val Val Gly
            340                 345                 350

Ile Gly Gly Thr Ala Gly Met Val His Pro Ser Thr Gly Tyr Met Val
        355                 360                 365

Ala Arg Thr Leu Ala Ala Ala Pro Ile Val Ala Asn Ala Ile Val Arg
    370                 375                 380

Tyr Leu Gly Ser Pro Ser Ser Asn Ser Leu Arg Gly Asp Gln Leu Ser
385                 390                 395                 400

Ala Glu Val Trp Arg Asp Leu Trp Pro Ile Glu Arg Arg Arg Gln Arg
                405                 410                 415

Glu Phe Phe Cys Phe Gly Met Asp Ile Leu Leu Lys Leu Asp Leu Asp
            420                 425                 430

Ala Thr Arg Arg Phe Phe Asp Ala Phe Phe Asp Leu Gln Pro His Tyr
        435                 440                 445

Trp His Gly Phe Leu Ser Ser Arg Leu Phe Leu Pro Glu Leu Leu Val
    450                 455                 460

Phe Gly Leu Ser Leu Phe Ser His Ala Ser Asn Thr Ser Arg Leu Glu
465                 470                 475                 480

Ile Met Thr Lys Gly Thr Val Pro Leu Ala Lys Met Ile Asn Asn Leu
                485                 490                 495
```

Val Gln Asp Arg Asp
500

<210> 24
<211> 526
<212> PRT
<213> Lycopersicon esculentum

<400> 24
Met Glu Cys Val Gly Val Gln Asn Val Gly Ala Met Ala Val Leu Thr
1 5 10 15

Arg Pro Arg Leu Asn Arg Trp Ser Gly Gly Glu Leu Cys Gln Glu Lys
20 25 30

Ser Ile Phe Leu Ala Tyr Glu Gln Tyr Glu Ser Lys Cys Asn Ser Ser
35 40 45

Ser Gly Ser Asp Ser Cys Val Val Asp Lys Glu Asp Phe Ala Asp Glu
50 55 60

Glu Asp Tyr Ile Lys Ala Gly Gly Ser Gln Leu Val Phe Val Gln Met
65 70 75 80

Gln Gln Lys Lys Asp Met Asp Gln Gln Ser Lys Leu Ser Asp Glu Leu
85 90 95

Arg Gln Ile Ser Ala Gly Gln Thr Val Leu Asp Leu Val Val Ile Gly
100 105 110

Cys Gly Pro Ala Gly Leu Ala Leu Ala Ala Glu Ser Ala Lys Leu Gly
115 120 125

Leu Asn Val Gly Leu Val Gly Pro Asp Leu Pro Phe Thr Asn Asn Tyr
130 135 140

Gly Val Trp Glu Asp Glu Phe Lys Asp Leu Gly Leu Gln Ala Cys Ile
145 150 155 160

Glu His Val Trp Arg Asp Thr Ile Val Tyr Leu Asp Asp Asp Glu Pro
165 170 175

Ile Leu Ile Gly Arg Ala Tyr Gly Arg Val Ser Arg His Phe Leu His
180 185 190

Glu Glu Leu Leu Lys Arg Cys Val Glu Ala Gly Val Leu Tyr Leu Asn
195 200 205

Ser Lys Val Asp Arg Ile Val Glu Ala Thr Asn Gly Gln Ser Leu Val
210 215 220

Glu Cys Glu Gly Asp Val Val Ile Pro Cys Arg Phe Val Thr Val Ala
225 230 235 240

Ser Gly Ala Ala Ser Gly Lys Phe Leu Gln Tyr Glu Leu Gly Ser Pro
245 250 255

Arg Val Ser Val Gln Thr Ala Tyr Gly Val Glu Val Glu Val Asp Asn
260 265 270

```
Asn Pro Phe Asp Pro Ser Leu Met Val Phe Met Asp Tyr Arg Asp Tyr
        275                 280                 285

Leu Arg His Asp Ala Gln Ser Leu Glu Ala Lys Tyr Pro Thr Phe Leu
    290                 295                 300

Tyr Ala Met Pro Met Ser Pro Thr Arg Val Phe Phe Glu Glu Thr Cys
305                 310                 315                 320

Leu Ala Ser Lys Asp Ala Met Pro Phe Asp Leu Leu Lys Lys Lys Leu
                325                 330                 335

Met Leu Arg Leu Asn Thr Leu Gly Val Arg Ile Lys Glu Ile Tyr Glu
            340                 345                 350

Glu Glu Trp Ser Tyr Ile Pro Val Gly Gly Ser Leu Pro Asn Thr Glu
        355                 360                 365

Gln Lys Thr Leu Ala Phe Gly Ala Ala Ala Ser Met Val His Pro Ala
    370                 375                 380

Thr Gly Tyr Ser Val Val Arg Ser Leu Ser Glu Ala Pro Lys Cys Ala
385                 390                 395                 400

Ser Val Leu Ala Asn Ile Leu Arg Gln His Tyr Ser Lys Asn Met Leu
                405                 410                 415

Thr Ser Ser Ser Ile Pro Ser Ile Ser Thr Gln Ala Trp Asn Thr Leu
            420                 425                 430

Trp Pro Gln Glu Arg Lys Arg Gln Arg Ser Phe Phe Leu Phe Gly Leu
        435                 440                 445

Ala Leu Ile Leu Gln Leu Asp Ile Glu Gly Ile Arg Ser Phe Phe Arg
    450                 455                 460

Ala Phe Phe Arg Val Pro Lys Trp Met Trp Gln Gly Phe Leu Gly Ser
465                 470                 475                 480

Ser Leu Ser Ser Ala Asp Leu Met Leu Phe Ala Phe Tyr Met Phe Ile
                485                 490                 495

Ile Ala Pro Asn Asp Met Arg Lys Gly Leu Ile Arg His Leu Leu Ser
            500                 505                 510

Asp Pro Thr Gly Ala Thr Leu Ile Arg Thr Tyr Leu Thr Phe
        515                 520                 525
```